# Investigating the Link Between Research Data and Impact

Phase II

Research Lead:
Dr Eric A. Jensen
eric@methodsinnovation.org

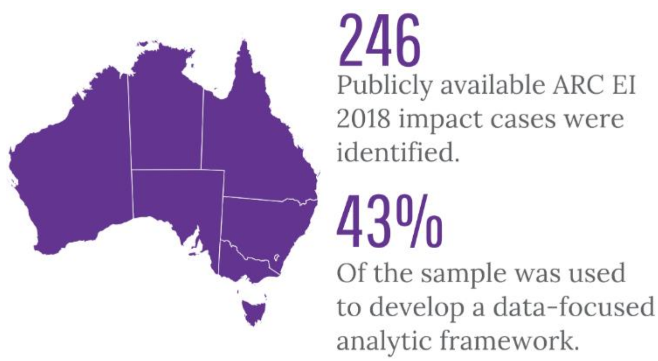
Lead Project Advisor:
Professor Mark Reed
mark.reed@ncl.ac.uk

# EXECUTIVE SUMMARY

The Institute for Methods Innovation – a research charity registered in the United States and United Kingdom – was commissioned by the Australian Research Data Commons (ARDC) to investigate how research data contributes to non-academic impacts by analysing existing case studies from the Australia Research Council (ARC) Engagement and Impact Assessment 2018. This represented a second phase to this work on the impacts of research data, with the first phase focusing on United Kingdom Research Excellence Framework (REF) impact case studies (ref.ac.uk).

## Project overview

The research involved analysing impact cases from the ARC's Engagement and Impact Assessment 2018. Only high scoring cases have currently been published by the ARC. These cases were sifted for the present research to focus our analysis on cases with an emphasis on 'data'. Relevant text segments from the published engagement and impact (E&I) case studies were extracted from the E&I case study documents. A content analysis was conducted on these data to identify patterns linking research data and impact. This analysis achieved a high level of scientific quality, based on established methodological standards.



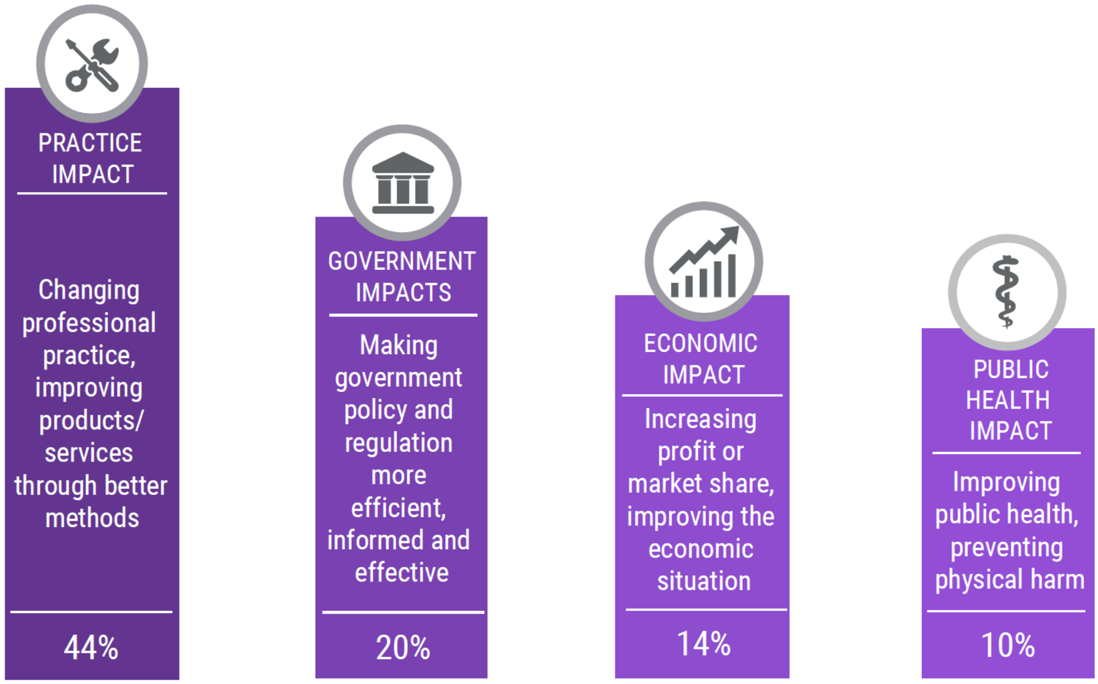

## What type of impact was developed from Australian research data?

The most prevalent type of research data-driven impact was *Practice impact* (44%). This category of impact includes changing the ways professionals operate and improving the quality of products or services through better methods, technologies, and responses to issues through better understanding. It also includes changing organisational culture and improving workplace productivity or outcomes.

*Government impacts* were the next most prevalent category identified in this research (20%). These impacts include the introduction of new policies and changes to existing policies, as well as reducing the cost to deliver government services, enhancing the effectiveness or efficiency of government services and operations, and more efficient to government planning.

Other relatively common types of research data-driven impacts were *Economic impact* (14%) and *Public Health Impact* (8%).

**What type of impact was developed from research data?**

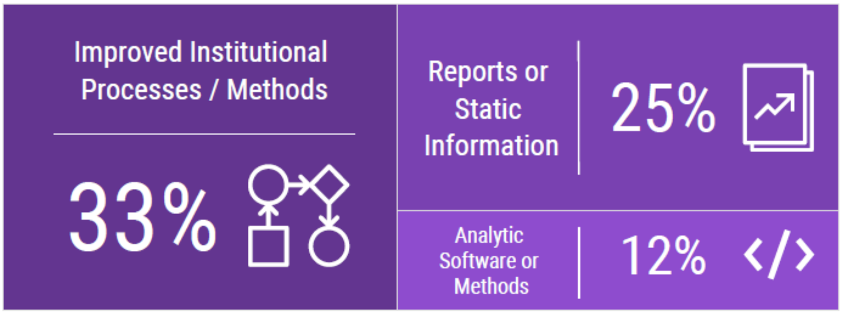



## How was impact developed from research data?

Impact from research data was developed most frequently through *Improved Institutional Processes / Methods* (33%). This relates to improving the way an institution operates, making it more efficient or effective at delivering outcomes. The second most common way of developing impact was via a *Report* (25%) of some kind; that is, the presentation of information based on the analysis and interpretation of relevant data. *Analytic Software or Methods* (12%) comprised the third most frequently used way of developing impact. Here, research data are used to generate or refine analytic software or methods which, in turn, generated impacts.

In general, research data itself rarely contributes directly to any impact. Instead, 99% of research data-linked impacts are indirectly associated with the identified impacts. Data need to be processed and conclusions or other value need to be drawn from them so that they can yield non-academic impacts.

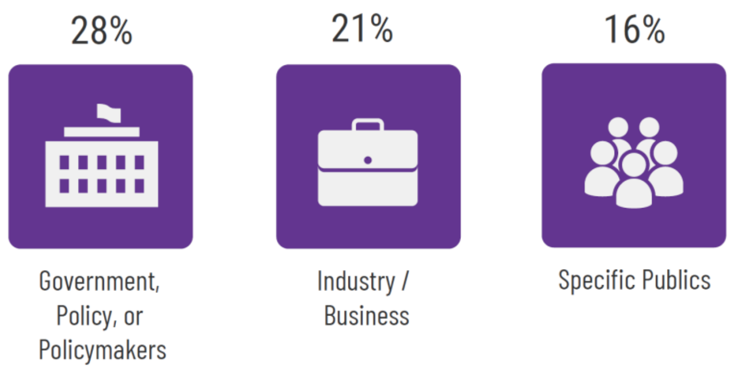
**How was data-linked impact developed?**




## Who benefited from the research data-linked impact?

*Government, Policy, or Policymakers* (28%), *Industry / Business* (21%), and *Specific Publics* (16%) were the most common types of beneficiaries from the research data-linked impacts we analysed. This finding is indicative of a two-step flow of research data-linked impact that ultimately reaches publics or wider non-academic stakeholders. Intermediaries such as the government, policymakers, and businesses are typically the primary beneficiaries of research data-based impacts. However, they often in turn use what they have gained to develop further insights, services, products, and policies that deliver broader public benefits.

**Who benefits from research data-linked impact?**




Looking at patterns in this analysis, the following statistically significant associations were identified:

- Reports or other types of static information tended to be used most with government stakeholders, and significantly less with industry / business.
- Analytic software or methods as well as shared technology or software were used more to develop impacts with industry / business, and significantly less with specific publics.
- Professionals as opposed to government stakeholders tended to benefit most from improved institutional processes.
- The general public tended to benefit from other impact instruments, e.g. gaining benefits via policy change.

## Where did the research data in these impact case studies come from?

We found that research data used in our sample of EI 2018 case studies were sourced from a range of different categories of people and organisations. *Specific Publics* (21%), such as hospital patients, were the largest category. We found that *Specific Publics* also tended to be beneficiaries of impacts related to data sourced from *Specific Publics*. Research data in our sample also commonly originated from *Industry / Business* (17%), and the *Natural Environment* (17%) and *Other Organisations* (17%).

Most of the impact-linked research data seemed to have been sourced by *Research Performing Organisations* (96%), such as universities. However, other contributing parties leading on primary sourcing and data collection may not have been included in the case study narratives.

## Conclusions

Various connections were found between predictor variables and certain outcomes such as identified types of impact and beneficiaries, for example:

- The *Field of Research*, that is, which type of academic discipline the research data-linked impacts were attributed to- was found to be a moderate predictor of impacts. For instance, practice impacts were significantly more commonly associated with the field of *Psychology and Cognitive Sciences* than with other research fields, and *specific publics* as research impact beneficiaries were disproportionately associated with the research field of *Aboriginal and Torres Strait Islander Research*
- On the other hand, whether the university associated with the impact-linked research data was part of the elite Group of Eight (Go8) overall did not or only weakly influenced impacts.

The analysis found that research data on their own rarely lead to impact, but instead they require analysis, curation, product development or other interventions to leverage broader non-academic value from the research data. These interventions help to bridge the gap between research data – which might otherwise go unused for the purpose of developing impact – and the diverse range of potential primary and secondary beneficiaries.

As such, impact from research data may be increased through closer links between government, industry and researchers, as well as capacity building at each of these levels. Capacity building initiatives can be aimed at potential impact beneficiaries, including supporting them to access useful sources of research data, and either understand and make use of this data or adapt it to serve new purposes. As such, the way that research data are made available, and the nature of the support available for interpreting and using this data, can affect how feasible it is to use that research data to develop new and creative pathways to impact.

Finally, there were strikingly high 'uniqueness' scores for the impacts linked to research data (93%), suggesting that most of the research data-linked impacts may have only been possible to develop through research data. However, limitations inherent in impact case studies have to be taken into account before drawing firm conclusions on this point.

Finally, it is worth noting that the findings from the Australian EI 2018 case studies are broadly similar to the UK impact case study findings from Phase I of this work. This similarity suggests that there may be structurally parallel patterns internationally in how research data are used to develop non-academic impacts.

# TABLE OF CONTENTS

# 1 INTRODUCTION

Comprising Phase II of the research conducted for the Australian Research Data Commons (ARDC) about the role of research data in developing impact, this report highlights key trends and patterns evident in the ARC Engagement and Impact 2018 case studies. Phase I of this project focused on the UK's REF impact cases.

Initially limited to high scoring cases by the ARC's selective publication of E&I impact cases, keywords (e.g. 'data' and 'database') were used to search for full impact narratives with high relevance to the present research within the ARC Engagement and Impact 2018 impact case database. All 246 publicly available impact cases were retrieved on 4 July 2019.

For the analysis, a subset of the sample (105 explicitly data-focused cases; approx. 43%) was used to develop an analytic framework, with specific categories, definitions, and examples. The framework was built through manually assessing this subset, extracting relevant parts, and refining it over the course of the analysis. This process provided key operational definitions and was designed to address the research questions for the project. In this Phase II research, three new dimensions were introduced into the analysis to assess the origins and sourcing of the research data, and how directly the data contributed to the impacts.

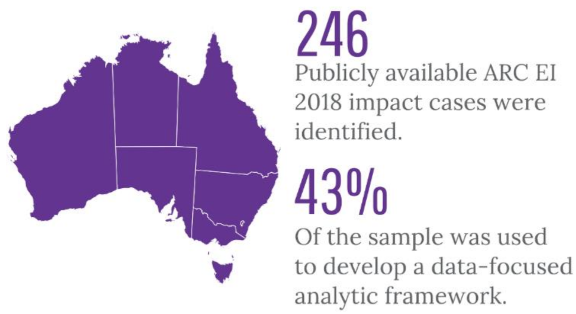



*Figure 1. Research data-linked impact dataset*

All categories were applied systematically through a well-established social research method known as 'content analysis'. This yielded quantified results, underpinned by intercoder reliability checks on randomly selected subsets of the data. To focus the analysis, relevant text passages from the engagement and impact case studies were identified first, followed by the main analysis.

This report begins with the results, showing what types of impact were most frequently associated with research data. The next section develops a portrait of the ways that impact was developed using research data. Then, we analyse the impact pathways associated with research data. After exploring the provenance and sourcing of research data in this dataset, we analyse whether research data offer a unique pathway to impact. Finally, these findings are placed in context with conclusions and recommendations, which align with the Phase I research.

The methods used for this study, including sampling, intercoder reliability analysis, and descriptive and inferential statistics can be found under Appendix A: methods.

## 1.1 Field of Research as a Predictor of Impact Patterns

Used as a 'predictor' variable in this analysis, the *Field of Research* (FoR) attributed to each impact case describes which research fields or academic disciplines are linked to the reported impact analysed in this study. This FoR information came with the dataset downloaded from ARC. The analysis assessed whether field of research influenced outcomes such as *Impact Category*, *Impact Instrument*, etc.

Table 1 shows how frequent each FoR category[1] was in the sample used for this analysis. The four most prevalent categories in our sample are *Medical and Health Sciences (Public and Allied Health Sciences)* (9%), *Engineering* (9%), *Agricultural and Veterinary Sciences* (8%), as well as *Environmental Sciences* (7%).

*Table 1. Prevalence of different Fields of Research in Phase II research sample*

| Field of Research | Frequency | Percent |
|---|---|---|
| Medical and Health Sciences (Public and Allied Health Sciences) | 67 | 9% |
| Engineering | 65 | 9% |
| Agricultural and Veterinary Sciences | 57 | 8% |
| Environmental Sciences | 49 | 7% |
| Commerce, Management, Tourism and Services | 47 | 7% |
| Law and Legal Studies | 45 | 6% |
| Psychology and Cognitive Sciences | 38 | 5% |
| Economics | 35 | 5% |
| Education | 35 | 5% |
| Interdisciplinary Research | 32 | 4% |
| Studies in Human Society | 32 | 4% |
| Studies in Creative Arts and Writing | 27 | 4% |
| Aboriginal and Torres Strait Islander Research | 26 | 4% |
| Biological Sciences | 26 | 4% |
| Built Environment and Design | 25 | 3% |

[1] In this report we are generally referring to the count among all instances of a variable (allowing for multiple per case) instead of the count of cases to which a category of a variable applied.

| | | |
|---|---|---|
| History and Archaeology | 22 | 3% |
| Medical and Health Sciences (Biomedical and Clinical Sciences) | 21 | 3% |
| Earth Sciences | 20 | 3% |
| Mathematical Sciences | 17 | 2% |
| Physical Sciences | 14 | 2% |
| Chemical Sciences | 4 | 1% |
| Language, Communication and Culture | 4 | 1% |
| Philosophy and Religious Studies | 4 | 1% |
| Information and Computing Sciences | 3 | 0% |

## 1.2 Group of 8 (Go8) status as a predictor of impact patterns

The second contextual factor we explored was membership in the Group of 8 (Go8) elite Australian universities for the institution submitting a case study. We investigated whether there is an association between Go8 university status and outcomes is the Go8-status[2]. Table 2 shows the distribution of Go8/non-Go8 universities in the sample.

*Table 2. Prevalence of Go8/non-Go8 universities in our sample of EI 2018 case studies*

| Go8 Status | Frequency | Percent |
|---|---|---|
| Go8 | 250 | 35% |
| Non-Go8 | 465 | 65% |

[2] Submitted EI 2018 case studies published on the ARC website included the submitting institution in the publicly available meta-data. This information was matched to the publicly available list of Go8 institutions to constitute this variable.

# 2 TYPES OF IMPACT ASSOCIATED WITH RESEARCH DATA

The first level of analysis in this research was designed to identify the types of impacts that are most commonly linked to research data. To ensure comparability with the Phase I results, the UK REF definition of impact has been retained: It defines impact as any positive effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia. This definition aligns closely with the ARC Engagement and Impact Assessment 2018 definition of impact: "Research impact is the contribution that research makes to the economy, society, environment or culture, beyond the contribution to academic research".

This section begins by defining the different types of impact that emerged from this analysis and explores the overall patterns that have been revealed by the data analysis.

## TYPES OF RESEARCH DATA-LINKED IMPACT IDENTIFIED

We begin by looking at the different types of impact associated with research data. The definitions for the different types of impact[3] are presented in Table 3.

*Table 3. Identified types of impact*

| Impact | Description |
|---|---|
| Government Spending / Efficiency Impact | Reducing the cost of delivering government services; increasing impact/quality of government service without raising cost. |
| Other Government / Policy Impact | Changing public policy or government regulations, or how either of these are implemented. |
| Practice Impact | Changing the ways that professionals operate; changing organizational culture; improving workplace productivity or outcomes; improving the quality of products or services through better methods, technology, understanding of the problems, etc. |
| General Public Awareness Impact | Improving public knowledge about a topic or increasing public visibility or attention for an issue. |
| Justice / Crime Reduction / Public Safety Impact | Reducing crime; Increasing efficiency in reducing crime; Improving justice outcomes (i.e. fairer; less cost; better social outcomes). |
| Public Health Impact | Improvements to the health of the population or a part of the population. |
| Economic Impact | Improvements to the economy or overall financial/economic situation. |
| Environmental Impact | Improvements in the natural environment, or reductions in threats or harm. |
| Other Kind of General Public Impact | Benefits for the general public (not professionals/government) that are not explicitly stated above in another category. |

[3] These categories were first identified and defined in an analysis of REF impact cases that were associated with research data for Phase I of this project.

| | |
|---|---|
| Other Non-Academic Impact | REF-eligible non-academic impacts not falling into any of the categories above. That is, cannot include academic publications or improvements to the teaching within a researcher's own institution. |
| Unclear / Uncertain | Not enough detail or clarity to clearly identify. |

## 2.1 Prevalence of Different Impact Types

The most prevalent[4] types of research data-driven impact in our sample were related to *Practice* (44%) and *Government* (20%), which includes both *Government Spending / Efficiency* (4%) and *Other Government / Policy* impacts (16%).

Likewise, other types of research data-driven impacts such as *Economic* impact (14%) and *Public Health* impact (8%) were also represented in a noteworthy minority of cases.

*Table 4. Prevalence of different types of impact associated with research data*

| Impact Type | Percentage |
|---|---|
| Practice Impact | 44% |
| Other Government / Policy Impact | 16% |
| Economic Impact | 14% |
| Public Health Impact | 8% |
| Other Kind of General Public Impact | 4% |
| Government Spending / Efficiency Impact | 4% |
| Environment Impact | 3% |
| General Public Awareness Impact | 3% |
| Justice / Crime Reduction / Public Safety Impact | 2% |
| Other Non-Academic Impact | 1% |

### Practice Impacts

The findings show that 44% of research data-linked impacts focused on practice. In these cases, the research data have been used to develop changes in the ways that professionals operate. These changes have a direct (or indirect) impact on outcomes such as organisational culture, workplace productivity or quality of products or service delivery through better methods, technology or understanding of the problems.

For example, as a result of the application of research data collected from a department store chain's customer loyalty programme, there was a marked improvement in the way marketing was approached and resources used:

[4] ARC EI 2018 impact case content with impact dimensions that fit in more than one field were categorised for each impact separately.

The research data were used to “develop a sophisticated advertising response model” with which the department store chain “was able to more efficiently allocate marketing resources across a range of media, and compare the returns of investing in new media, such as the internet and social media, as opposed to traditional media, such as television and newspapers”. This example shows research data being used to change how a business operated in order to improve efficiency and business outcomes.

### Government and Policy Impacts

16% of impacts were related to one of the following developments:

- Introducing new public policy or changing existing policy
- Increasing impact/quality of government services, without raising costs
- Reducing the cost of delivering government services

An example of a case that focuses on changing policy or government regulations involved the then Department of Health and Ageing. Here, research on the Extended Medicare Safety Net (EMSN) resulted in reviews which led to “reforms […] contained in the Health Insurance (EMSN) Act which was enacted in Oct 2009 and came into effect in Jan 2010”.

### Economic Impact & Public Health Impact

Research data were also used to develop impacts for both the economy (14%) and public health (8%). Examples of economic impact included cases that delivered increased productivity and therefore financial competitiveness. Public health impacts included improved health outcomes or prevention of physical harm to the general population or specific publics, such as the use of flooding data to improve emergency preparedness and lower fatality rates.

## 2.2 Links between impact type and field of research

The Fields of Research (FoR) associated with research data-linked impacts listed in Table 1 had a weak to moderate influence over the type of impact generated.[5] Knowing the field of research, one can accurately predict more than one quarter (29%) of the variability in impact type in our sample.

### Practice Impact

Practice impacts were significantly more commonly associated with the field of *Psychology and Cognitive Sciences* than with other research fields.[6]

### Government Spending / Efficiency Impact

Research data-linked impacts relating to government spending and efficiency tended to stem from the research field of *Environmental Sciences*.[7] That is, in the present Phase II sample, research on the environment disproportionately leads to impacts that change how the government or government agencies target and improve value for their spending.

[5] $\chi^2(207, N = 715) = 538182$, $p < .001$, $V = .29$
[6] Expected: 16%, observed: 45%, difference: +29%, $p < .001$
[7] Expected: 4%, observed: 18%, difference: +14%, $p < .001$

### Justice / Crime Reduction / Public Safety Impact

*Justice / Crime Reduction / Public Safety Impacts* were significantly more often found in research linked to the field of *Commerce, Management, Tourism and Services* than in other fields of research.[8]

### Public Health Impact

Unsurprisingly, the research field of *Medical and Health Sciences* was significantly more likely than other fields to deliver public health impacts.[9] This result indicates that medical and health research is delivering on its expected societal value of effecting positive health outcomes for specific or broader publics.

## 2.3 Links between impact type and Go8 membership

We found no statistically significant differences between Go8 universities and other Australian universities in how often certain impact types were identified.[10]

[8] Expected: 2%, observed: 13%, difference: +11%, $p < .001$
[9] Expected: 8%, observed: 27%, difference: +19%, $p < .001$
[10] $\chi^2(9, N = 715) = 13589$, $p = .11$

# 3 How impact develops from research data

This section focusses on the ways that impact was leveraged from research data in the sample of EI 2018 case studies we analysed.

## 3.1 Different instruments to develop impact using research data

The impact is developed from research data in several different ways. Here, we analyse the nature of these different impact-generating instruments. The impact development approaches we identified are summarised in the table below.

*Table 5. Identified ways of developing impact*

| How impact was developed | Description of impact instrument |
|---|---|
| Searchable Database | A database that can be accessed to view the research data in a dynamic way (that is, offers the ability to select variables/filters, allowing for customised information to be accessed by users to use for their own purposes). |
| Reports or static information | Report containing pre-analysed/curated information, a static database, results tables or other methods of presenting the research data as processed information to be used without customisation or filtering of the data. |
| Mobile App | An application designed for smartphone or tablet to access the research data or an analysis/results of the data. |
| Analytic Software or Methods | Research data used to generate or refine software or research/analytic methods or statistical models. |
| Improved Institutional Processes / Methods | Research data used to make an institution's way of operating better/more efficient or more effective at delivering outcomes. |
| Sharing of Raw Data | Research data has an impact via being shared with others (in raw or minimally anonymised form) outside of the research team that generated the data so that they can do something with it (e.g. further analysis, etc.). |
| Sharing of Tech / Software | The research data have an impact via sharing technology or software that was created using the research data or that uses the research data somehow. |
| Other Impact Instrument | A clearly identifiable impact instrument that does not fit into any of the categories listed above. |
| Unclear / Uncertain | Impact instrument that is not detailed enough to clearly place into any pre-specified category. |

## 3.2 Prevalence of Impact Development Instruments

This section explores the frequency with which the different ways of developing impact appeared in the cases analysed. The most common[11] ways of developing impact from research data were *Improved Institutional Processes / Methods* (33%), *Reports / Static Information* (25%) and *Analytic Software or Methods* (12%).

*Table 6. Prevalence of Impact Development Instruments*

| How impact was developed | Percentage |
|---|---|
| Improved Institutional Processes / Methods | 33% |
| Report or Static Information | 25% |
| Analytic Software or Methods | 12% |
| Sharing of Tech / Software | 9% |
| Unclear / Uncertain | 8% |
| Other Impact Instrument | 7% |
| Sharing of Raw Data | 4% |
| Searchable Database | 2% |
| Mobile App | 1% |

### Institutional Processes or Methods

Improving 'institutional processes or methods' was a major instrument for developing impact identified in this study. One example of this category of impact development comes from research on pathogen survival during biosolid storage which provided evidence supporting the reduction of storage time and therefore cost:

> "[The research] has improved the assessment and management of wastewater risk. Their research has informed a reduction in biosolid stockpiling times from three to one years, producing substantial economic and environmental benefit. […] At the Boneo treatment plant, the deferred capital expenditure led to a saving of $1 million[, and r]educing stockpiling times to one year has addressed problems such as greenhouse gas emissions, environmental harm and issues with local amenity".

Other examples of how impacts were developed from research data via enhanced institutional processes include changes in medical practices resulting in improved medical services, as well as green enhancements in government practices with pro-environmental outcomes.

[11] As multiple ways of developing impact could be used in tandem, the analysis allowed for multiple impact instruments to be identified for a single impact.

### Reports or Static Information

Research data were often used to develop impact through the production of 'reports' or other similar types of prepared information. Such reporting distils research data in a way that makes them intelligible for institutions, making the data useful for a wider user base outside of academia.

### Analytic Software or Methods

'Analytic Software or Methods' were also used to develop impact. An example of this category of impact development can be drawn from research which led to the creation of methods for the assessment of mental illness treatments, and identifying emerging mental health issues:

> "Research has had a direct impact on health and well-being [by c]reating a standardised way of measuring whether particular types of music are helping or harming adolescents suffering from certain mental illnesses. […] [The] 'Healthy-Unhealthy Music Scale' (HUMS) [is] a 13-question tool for use in clinical practice [which is used] to identify emerging mental health issues".

## 3.3 Links between impact instrument and field of research

We found moderate statistical differences in the impact instruments used across different fields of research in the EI 2018 case studies we analysed.[12] Nearly one third (32%) of the differences in which impact instruments are used can be predicted by the field of research.

### Searchable Database

Research data-linked impacts generated by means of searchable databases were disproportionately associated with the research fields of *Aboriginal and Torres Strait Islander Research*[13] and *History and Archaeology*[14]*.* For example, research data were used to generate databases particular user communities could access to reconnect with ancestral history.

### Report or Static Information

Reports was significantly less common as an impact instrument in impact case studies associated with the fields of *Agricultural and Veterinary Science*[15] and *Engineering*[16] when compared with other FoR categories. Impacts associated with these two fields tend to be developed through impact instruments *other than* reports.

---

[12] $\chi^2(184, N = 906) = 738104$, $p < .001$, $V = .32$

[13] Expected: 2% observed: 17%, difference: +15%, $p < .001$

[14] Expected: 2%, observed: 20%, difference: +18%, $p < .001$

[15] Expected: 25%, observed: 1%, difference: -24%, $p < .001$

[16] Expected: 25%, observed: 3%, difference: -22%, $p < .001$

### Analytic Software or Methods

The fields of *Agricultural and Veterinary Sciences*[17], *Earth Sciences*[18] and *Mathematical Sciences*[19] used the impact instrument of analytic software or methods more than other impact-generating approaches. Impact case studies from these fields used research data to develop analytic software or methods that could improve outcomes in affiliated practical domains, such as boosting agricultural outcomes and efficiency.

### Improved Institutional Processes / Methods

The impact development approach of improving institutional processes/methods was significantly less prevalent in impact cases from the field of *Economics*[20] when compared to other FoR categories.

### Sharing of Raw Data

Sharing raw data was the only impact instrument disproportionately associated with the field of *Economics*[21] in our sample.

### Sharing of Technology / Software

The fields of *Agricultural and Veterinary Science*[22], *Engineering*[23], and *Medical and Health Sciences*[24] used technology/software as an impact instrument more than other FoR categories in the case studies we analysed.

## 3.4 Links between impact instrument and Go8 membership

Overall, there was a weak statistical association between Go8 membership and the type of impact instrument used.[25] In case studies submitted by Go8 universities, sharing technology/software was more frequently used than other impact instruments.[26] However, just 15% of the variability in impact instruments used in our sample can be predicted by Go8 membership.

[17] Expected: 12%, observed: 31%, difference: +19%, $p < .001$
[18] Expected: 12%, observed: 36%, difference: +24%, $p < .001$
[19] Expected: 12%, observed: 48%, difference: +36%, $p < .001$
[20] Expected: 33%, observed: 5%, difference: -28%, $p = .012$
[21] Expected: 4%, observed: 21%, difference: +17%, $p < .001$
[22] Expected: 9%, observed: 27%, difference: +18%, $p < .001$
[23] Expected: 9%, observed: 26%, difference: +17%, $p < .001$
[24] Expected: 9%, observed: 43%, difference: +34%, $p < .001$
[25] $\chi^2(8, N = 906) = 19651$, $p = .012$, $V = .15$
[26] Expected: 9%, observed: 13%, difference: +4%, $p = .033$

# 4 Direct versus indirect impact pathways

The next section is concerned with whether the research data-linked impacts were a direct result of the research data. Specifically, the analysts for this research assessed whether the impact was generated through direct engagement with the research data or an indirect connection with the data, for instance through other products or services.

## 4.1 Analysing directness of impact pathways

It was found that only about 1% of the impact pathways identified in the sample were connected directly to the research data, without any intervening steps or instruments standing between the research data and the impact. This indicates that most research data need at least some form of analysis or other type of processing (e.g. a research report) to deliver an impact.

*Table 7. Prevalence of indirect versus direct research data-linked impact pathways*

| Impact Pathway | Frequency | Percentage |
| --- | --- | --- |
| **Indirect** | 707 | 99% |
| **Direct** | 6 | 1% |

## 4.2 Links between directness of impact pathway and field of research

There was a moderately strong link between field of research and the directness of the impact pathway overall, with 27% of the variance in the directness of the impact pathway predicted by the field of research.[27]

## 4.3 Links between impact pathway and Go8 membership

We found no statistically significant differences in the directness of impact pathways between Go8 universities and other Australian (non-Go8) universities in this sample.[28]

[27] $\chi^2(46, N = 715) = 100665, p < .001, V = .27$
[28] $\chi^2(2, N = 715) = 13159, p = .37$

# 5 THE BENEFICIARIES OF RESEARCH DATA-LINKED IMPACT

This section focuses on who is benefiting from research data-linked impacts and related impact development patterns. Here, we report on the people and organisations that gained value from the research data-linked impacts in our sample.

## 5.1 DEFINING IMPACT BENEFICIARY CATEGORIES

The table below contains definitions of the different impact beneficiary categories which were retained from the Phase I research.

*Table 8. Definitions of Impact Beneficiary Categories*

| Beneficiary | Description |
|---|---|
| General Public | Unspecified public as beneficiary; or relatively undifferentiated, society-wide, community-wide or a national or regional audience. |
| Specific Public | This category applies when a particular demographic or psychographic category of non-professional/non-governmental/non-business beneficiary is specified (e.g. children/mothers/art museum visitors/etc.). |
| Media | Research data improving media services, enabling data journalism, resulting in news coverage or news insights, documentaries, or entertainment media. |
| Professionals | Improved capacities, skills, employment options, increased salaries or benefits, greater influence, etc. |
| Government, Policy, or Policymakers | All levels of government (e.g. affecting how government delivers services, prioritises etc) or government policy (e.g. work cited or underpinning legislation, regulation or new policy initiatives) or policymaking (e.g. expert committees or feeding into city or government department research designed to inform policy or regulations). |
| Industry / Business | Improving business outcomes, profits, service/product quality, reducing environmental impact, etc. |
| Other Organization | Non-governmental / Non-business organization not covered by the above categories. |
| Natural Environment | Improved environmental outcomes, or reduction in damage/threat. |
| Unclear / Uncertain | Not enough detail provided to clearly identify the nature of the beneficiary. |

## 5.2 Research data-linked impact beneficiaries

This section addresses how common[29] the impact beneficiary categories were in our sample of EI 2018 case studies.

*Table 9. Prevalence of different types of beneficiaries in the analysed impact cases*

| Impact beneficiaries[30] | Percentage |
|---|---|
| Government, Policy, or Policymakers | 28% |
| Industry / Business | 21% |
| Specific Public | 16% |
| Other Organization | 13% |
| General Public | 8% |
| Professionals | 7% |
| Natural Environment | 5% |
| Media | 1% |

*Government, Policy, or Policymakers* were the primary beneficiaries in 28% of the identified research data-linked impacts in this study. This category of impact beneficiary encompasses all levels of government, including improvements of government programmes and policy change. This category of beneficiary often gained value from research input, which informed policy introduction or change. It also benefited from research data-informed interventions to improve efficiency or effectiveness of government programmes or services. These improvements, in turn, may be traceable to broader public benefits.

*Industry / Business* comprised the second most frequent (21%) category of impact beneficiaries, gaining improved business outcomes, profits, enhanced services and product quality and other benefits.

*Specific Public* was also a relatively frequent category of beneficiary from research data-driven impact (16%), comprised of particular groups of people receiving improved products or services specific to their needs. This category also included underprivileged publics as beneficiaries of empowerment and increased equity.

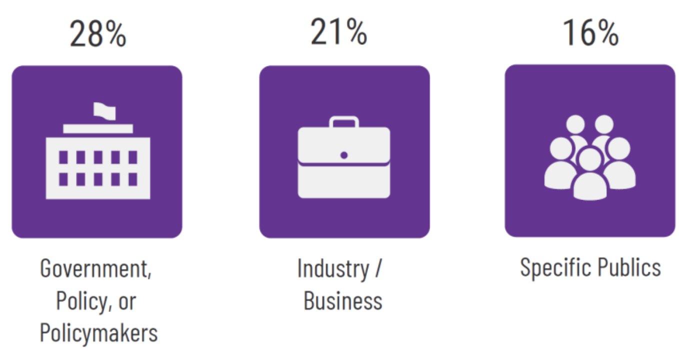



*Figure 2. Three most prevalent types of beneficiaries in the analysed impact cases*

[29] More than one beneficiary would be identified for each impact during the data analysis.
[30] While not of primary interest in this analysis, the 'Unclear / Uncertain Beneficiary' category was represented in 2% of identified research data-linked impacts.

An example of benefits delivered to *Specific Publics* comes from research conducted at the Curtin University of Technology: survey data and a subsequent clinical trial led to improved “skin tear prevention and management outcomes for elderly Western Australians” (i.e. ‘*elderly* Western Australians’ are the ‘specific public’ in this case). This kind of specificity in development of public benefit may be tied to the fact that research data are so often grounded in particular populations.

### 5.2.1 The link between how impact is developed and who it benefits

The following includes only variables with statistically significant. The differences in marginal percentages indicate either positive or negative associations. In general, the analysis revealed a weak association between the Impact Instruments and Beneficiaries.[31] The following points summarise statistically significant patterns identified through correlation analysis:

- Reports or other types of static information tended to be used with government or policy stakeholders[32], and significantly less with industry / business[33].
- Analytic software or methods[34] as well as shared technology or software[35] were used more to develop impacts with industry / business, and significantly less with specific publics[36].
- Professionals[37] as opposed to government stakeholders[38] tended to benefit from improved institutional processes.
- The general public tended to benefit from other impact instruments[39], e.g. gaining directly from policy change.

## 5.3 Links between beneficiaries and field of research

We found a moderate influence of the field of research over the data-linked impact beneficiaries in our sample.[40] Here, the field of research can accurately predict one third (33%) of the variability in impact beneficiaries.

### Specific Public

Research data-linked impacts benefitting *specific publics* were disproportionately associated with the research field of *Aboriginal and Torres Strait Islander Research* in this dataset.[41] This is unsurprising given that issues within the specific public of indigenous community groups were often the focus of the research data. The impacts delivered from these research initiatives ultimately benefited these same groups of people.

---

[31] $\chi^2(64, N = 1170) = 299855$, $p < .001$, $V = .18$
[32] Expected: 27%, observed: 41%, difference: +14%, $p < .001$
[33] Expected: 24%, observed: 12%, difference: -12%, $p < .001$
[34] Expected: 24%, observed: 42%, difference: +18%, $p < .001$
[35] Expected: 24%, observed: 58%, difference: +34%, $p < .001$
[36] Expected: 15%, observed: 5%, difference: -10%, $p = .049$
[37] Expected: 6%, observed: 11%, difference: +5%, $p < .001$
[38] Expected: 27%, observed: 20%, difference: -7%, $p = .002$
[39] Expected: 8%, observed: 26%, difference: +18%, $p < .001$
[40] $\chi^2(184, N = 922) = 837565$, $p < .001$, $V = .33$
[41] Expected: 16%, observed: 39%, difference: +23%, $p = .008$

### Professionals

This category was more commonly associated with the field of *Education* than other FoR categories.[42] This indicates that research on education benefitted professionals such as teachers and others working in the education sector.

### Government, Policy, or Policymakers

In the field of *Agricultural and Veterinary Sciences,* impacts benefitting government/policy stakeholders were significantly underrepresented when compared to other fields of research.[43] However, government/policy stakeholders were more likely to be beneficiaries of impacts associated with *Economics*[44] and *Environmental Sciences*[45] than other FoR categories. Many of the government and policy impacts emerged from economics and environmental research data.

### Industry / Business

Industry and business stakeholders were more frequent beneficiaries of research data-linked impacts stemming from *Agricultural and Veterinary Sciences*[46], *Biological Sciences*[47] and *Engineering*[48] than other research fields. On the other hand, the fields of *Education*[49] and *Medical and Health Sciences*[50] were comparatively less likely to deliver impacts that benefitted industry and business stakeholders.

### Other Organisation

The fields of *Education*[51], *History and Archaeology*[52], and *Medical and Health Sciences*[53] tended to report impacts that benefitted 'other organisations' such as hospitals and clinics, schools, and museums at a greater frequency than other FoR categories.

### Natural Environment

The natural environment as a beneficiary was associated with impacts emerging from *Interdisciplinary Research*[54] and *Mathematical Sciences*[55] more than other FoR categories. Also, data from *Interdisciplinary Research* often resulted in a broader variety of impacts through different impact instruments, meaning that the natural environment could be a frequent co-beneficiary.

[42] Expected: 7%, observed: 20%, difference: +13%, $p = .025$
[43] Expected: 28%, observed: 0%, difference: -28%, $p < .001$
[44] Expected: 28%, observed: 54%, difference: +26%, $p = .03$
[45] Expected: 28%, observed: 62%, difference: +34%, $p < .001$
[46] Expected: 21%, observed: 75%, difference: +54%, $p < .001$
[47] Expected: 21%, observed: 64%, difference: +43%, $p < .001$
[48] Expected: 21%, observed: 51%, difference: +30%, $p < .001$
[49] Expected: 21%, observed: 0%, difference: -21%, $p = .033$
[50] Expected: 21%, observed: 2%, difference: -19%, $p < .001$
[51] Expected: 13%, observed: 32%, difference: +19%, $p = .004$
[52] Expected: 13%, observed: 38%, difference: +25%, $p = .012$
[53] Expected: 13%, observed: 29%, difference: +16%, $p < .001$
[54] Expected: 5%, observed: 18%, difference: +13%, $p = .03$
[55] Expected: 5%, observed: 32%, difference: +27%, $p < .001$

## 5.4 Links between beneficiaries and Go8 membership

We found no statistically significant differences between beneficiaries of cases submitted by universities of the Go8 and those submitted by other Australian universities.[56]

[56] $\chi^2(8, N = 922) = 13159, p = .11$

# 6 Uniqueness of impact pathways

This section reports on findings regarding the uniqueness of impact pathways offered by research data. The analysts assessed, based on the information available within the impact case study, whether the identified impact could have developed *without* the underlying research data. That is, was the research data required for the impact to occur?

## 6.1 The impact counter-factual analysis

Using the content of the Engagement and Impact case study narratives, the analysts answered the following question:

*Based on the information in the case study, was the research data required for the identified impact to exist?*

The systematic application of this impact counter-factual question to the case studies in our sample revealed a striking finding: Analysts judged that 93% of identified impacts in the EI 2018 case studies we analysed would not have been realised without research data. That is, these identified impacts could not have been developed without the supporting research data.

*Table 10. Prevalence of impacts that exclusively required research data to be developed*

| | Frequency | Percentage |
|---|---|---|
| **Impact required research data** | 666 | 93% |

This finding is particularly noteworthy because the full set of available Australian E&I case studies was included in this analysis.

## 6.2 Examples of unique research data-based impact pathways

To illustrate how research data can provide a unique pathway to achieving an identified impact, two examples of impacts that were judged to rely exclusively on research data are presented below.

### Example 1

Research from the University of Western Australia which “involved an analysis of hundreds of millions of transitions, quotes and orders in markets trading over US$30 trillion annually” revealed “a price bias and an exploitable trade advantage in the 100 year old mechanism used to set the daily benchmark price of precious metals globally”. This directly led to media coverage, increased public awareness, and subsequent justice outcomes through legal processes. The data on which the research was based was essential for these impacts to occur.

Example 2

An impact case focused on research in the area of “ethical, social and legal implications of biotechnologies”. The Centre for Law and Genetics (CLG) “research has driven change to IP law. The CLG’s IP research began in 2002 with an empirical study of the Australian medical biotech industry, involving participants from universities, biotech companies and diagnostics facilities. […] This led to the Australian Law Reform Commission (ALRC) engaging CLG Director Professor Dianne Nicol to contribute to its 2004 report (No.99) on Gene Patenting and Human Health. The ALRC review drew heavily on CLG research (140+ references). The 2003 empirical study was referenced in the Explanatory Memorandum to the Intellectual Property Laws Amendment (Raising the Bar) Act 2012 (Cth).”

The study’s empirical research directly informed, and perhaps even sparked, the ALRC review. Additionally, it influenced changes to intellectual property law and recommendations based on the research were adopted in the legal reform. This comprises a further example of an impact pathway that required research data as a necessary dimension.

## 6.3 Links between impact counter-factual and field of research

A weak but statistically significant association between the fields of research and the uniqueness of the research data-based impact pathway (i.e. ‘impact counter-factual’) was found.[57] Less than a quarter (23%) of the differences in the impact counter-factual can be explained by the field of research.

## 6.4 Links between impact counter-factual and Go8 membership

We did not find any statistically significant differences in the impact counter-factual variable between universities of the Go8 and other Australian universities.[58]

## 6.5 Limitations

The high percentage of ‘unique’ impact pathways identified in this analysis calls for re-visiting the limitations of the present analysis. Here, we are relying on impact case narratives that are being crafted to tell stories about the impact of research. There may be an incentive for case study authors to emphasise how essential the research (and also research data) were to the impacts described in the case. We must be cautious, therefore, in making generalizations to all research linked impacts that may be taking place, many of which may have been systematically omitted from EI 2018 case studies specifically because of concern about multi-dimensional pathways to impact making causal claims more tenuous and less likely to receive high scores.

[57] $\chi^2(69, N = 715) = 117415, p < .001, V = .23$

[58] $\chi^2(3, N = 715) = 4361, p = .23$

# 7 Provenance of the Research Data Used to Generate Impact

In this section, the data provenance was assessed, i.e. where the research data was drawn from by the collecting entity. Based on judgment and the information available, the analysts determined the original source of the data linked to an impact.

## 7.1 Defining Different Research Data Provenance Categories

The table below contains definitions of the different research data origins we developed for this analysis. These categories too were retained from the Phase I work.

*Table 11. Identified categories for research data provenance*

| Data provenance | Description |
|---|---|
| General Public | Unspecified public/demographic category as data provenance; or society-wide, community-wide or a national/regional audience. |
| Specific Public | Specified public/demographic category of e.g. customers/children/mothers/art museum visitors/ etc. |
| Media | Media services, (data) journalism, press releases, news coverage or news insights, documentaries or entertainment media as data provenance. |
| Professionals | Professionals with specific capacities, skills, employment options, increased salaries or benefits, greater influence, etc. as data provenance. |
| Government, Policy, or Policymakers | All levels of government (e.g. affecting how government delivers services, prioritizes etc.) or government policy (e.g. work cited or underpinning legislation, regulation or new policy initiatives) or policymaking (e.g. expert committees or feeding into city or government department research designed to inform policy or regulations). |
| Industry / Business | Businesses/industry, business outcomes, profits, services/products or similar as data provenance. |
| Other Organization | Non-governmental / Non-business organization not covered by the above categories. |
| Natural Environment | Environmental / natural data provenance. |
| Unclear / Uncertain | Not enough detail provided to clearly identify the data provenance. |

## 7.2 Impact-linked data provenance analysis

This section shows the overall results for the categories of research data provenance we identified in our sample of EI 2018 case studies.

*Table 12. Prevalence of different types of data provenance*

| Data provenance | Percentage |
|---|---|
| Specific Public | 21% |
| Industry / Business | 17% |
| Natural Environment | 17% |
| Other Organisation[59] | 17% |
| General Public | 13% |
| Unclear / Uncertain | 9% |
| Government, Policy, or Policymakers | 4% |
| Professionals | 3% |

The most common origin of impact-linked research data was *specific publics* (21%), because research presented in these impact case studies often focussed on certain groups of people such as school children, women in the workplace, or patients with certain diseases or disorders. Indeed, we found a pattern in which *Specific Publics* are often both the Data Source and the Beneficiary in the same cases[60], indicating that impacts from research data that have been sourced from specific publics also tend to benefit those same categories of people.

For example, research on Fetal Alcohol Spectrum Disorder (FASD) in Aboriginal communities benefited members of those communities: "The Lililwan study […] contributed to a 20% decrease in alcohol use in pregnancy in the Fitzroy Valley, improved child health, and behavioural change. […] The data collected enabled advocacy for better diagnosis, treatment and education".

Industry / business (17%) and the natural environment (17%) were approximately equally frequent sources of data. Here, again we found a relationship between the data provenance and beneficiary categories, with data sourced from industry/business tending to in turn be used to develop impacts that benefit industry/business (and the same for the natural environment).[61][62] For example, in one case, data collected from a sewage treatment plant led to more efficient biosolid stockpiling processes that, in turn, financially benefitted that industry. Likewise, environmental data about regional biodiversity and vegetation was used to led to environmental policy reforms and strategies that improved conservation.

[59] The data provenance category of *Other Organisation* includes different (non-industry, non-governmental) types of organisations such as universities and hospitals, where data was generated which mostly had impacts benefitting publics, the government, industry, and other entities.
[60] Expected: 16%, observed: 30%, difference: +14%, $p < .001$
[61] Expected: 21%, observed: 46%, difference: +25%, $p < .001$
[62] Expected: 5%, observed: 19%, difference: +14%, $p < .001$

## 7.3 Links between data provenance and field of research

We found that different fields of research were strongly associated with particular categories of data sourcing.[63] Over half (54%) of the variability in Data Provenance can be accurately predicted by the Field of Research in our sample.

### General Public

The general public were disproportionately the source of research data for the research fields of *Built Environment and Design*[64], *Economics*[65], and *Studies in Human Society*[66] when compared to other Fields of Research.

### Specific public

The *Specific public* category was significantly less common as a data source for the fields of *Agricultural and Veterinary Sciences*[67], and *Engineering*[68]. However, specific publics tended to be the origin of data for impacts resulting from *Medical and Health Sciences*[69], and *Psychology and Cognitive Sciences*[70].

### Professionals

Professionals were more likely to be the source of data for impacts associated with the fields of *Commerce, Management, Tourism and Services*[71], and *Education*[72] in our sample.

### Government, Policy, or Policymakers

Impacts developed from *Economics* research tended to originate from government data – specifically government data related to the economy and business.[73]

### Industry / Business

Industry / business was less likely to be a data source for impacts associated with *Medical and Health Sciences*[74]. However, the category of industry/business was more likely to be associated with the fields of *Agricultural and Veterinary Sciences*[75], *Commerce, Management, Tourism and Services*[76], and *Engineering*[77] than other fields in our sample.

---

[63] $\chi^2(161, N = 715) = 1469093$, $p < .001$, $V = .54$
[64] Expected: 13%, observed: 52%, difference: +39%, $p < .001$
[65] Expected: 13%, observed: 40%, difference: +27%, $p < .001$
[66] Expected: 13%, observed: 47%, difference: +34%, $p < .001$
[67] Expected: 21%, observed: 0%, difference: -21%, $p = .008$
[68] Expected: 21%, observed: 0%, difference: -21%, $p < .001$
[69] Expected: 21%, observed: 52%, difference: +31%, $p < .001$
[70] Expected: 21%, observed: 61%, difference: +40%, $p < .001$
[71] Expected: 3%, observed: 19%, difference: +16%, $p < .001$
[72] Expected: 3%, observed: 14%, difference: +11%, $p = .025$
[73] Expected: 4%, observed: 43%, difference: +39%, $p < .001$
[74] Expected: 17%, observed: 0%, difference: -17%, $p = .024$
[75] Expected: 17%, observed: 42%, difference: +25%, $p < .001$
[76] Expected: 17%, observed: 40%, difference: +23%, $p < .001$
[77] Expected: 17%, observed: 52%, difference: +35%, $p < .001$

### Natural Environment

Data originating from the natural environment was disproportionately likely to be associated with impacts emerging from *Chemical Sciences*[78], *Earth Sciences*[79], *Environmental Sciences*[80], *Interdisciplinary Research*[81], and *Mathematical Sciences*[82]. Natural environment data were, however, significantly less likely to be associated with impacts linked to the *Medical and Health Sciences*[83] when compared to other fields of research.

## 7.4 Links between data provenance and Go8 membership

We found no statistically significant differences in research data provenance in cases submitted by Go8-universities versus other (non-Go8) Australian universities.[84]

[78] Expected: 17%, observed: 100%, difference: +83%, $p = .001$
[79] Expected: 17%, observed: 65%, difference: +48%, $p < .001$
[80] Expected: 17%, observed: 80%, difference: +63%, $p < .001$
[81] Expected: 17%, observed: 56%, difference: +39%, $p < .001$
[82] Expected: 17%, observed: 65%, difference: 48%, $p < .001$
[83] Expected: 17%, observed: 0%, difference: -17%, $p = .024$
[84] $\chi^2(7, N = 715) = 11640$, $p = .11$

# 8 Sourcing of the Impact-Linked Research Data

Different than data provenance, this section deals with who sourced the data. The analysts assessed this based on the information available within each impact case study.

## 8.1 Defining Different Data-Sourcing Organisations

The table below contains definitions of research data sourcing included in this analysis. The categories were adopted from Phase I to ensure comparability.

*Table 13. Identified research data-sourcing organisations featured in EI 2018 cases in Phase II of this research*

| Data source | Description |
|---|---|
| Research Performing Organisation (RPO) | Organisation performing research or housing research activities, such as universities, institutes, academies, or similar. |
| Research Funding Organisation (RFO) | Any organization funding research and sourcing or collecting data. |
| Non-Governmental Organisation (NGO) | Non-Governmental Organizations, such as environmental or health organizations, associations or other citizen founded organizations. |
| Quasi-Governmental Organisation | A corporation or body with a public mandate that is directly supported by the government. Also known as 'arms-length' bodies that operate independently from the government. Quasi-Governmental Organizations are usually involved in providing oversight, funding, or have accountability for public benefit. |
| Government | All levels of government, public services, or policymaking (e.g. expert committees, government departments or other governmental entities). |
| Industry / Business | Businesses / industry, small companies, or for-profit organizations providing services or products. |
| Other data sources | Any other data sources, which can be specifically identified, but are not covered by any of the above options. |
| Unclear / Uncertain | Not enough detail provided to clearly identify the data source. |

## 8.2 Data sourcing analysis

Here, we show how common the different data sourcing/collecting categories are within the identified research data-linked impacts in our sample.

*Table 14. Prevalence of different types of data sourcing organisations*

| Data sourcing organisation | Percentage |
|---|---|
| Research Performing Organisation (RPO) | 96% |
| Industry / Business | 2% |
| Quasi-Governmental Organisation | 1% |
| Government | 1% |
| Non-Governmental Organisation (NGO) | 1% |
| Unclear / Uncertain | 0% |

The overwhelming majority of the research data from the EI 2018 impact cases was sourced by Research Performing Organisations (96%), because most research was conducted by university research teams. However, other contributing parties leading on data collection may not have been included in the case study narratives.

Indeed, there may be a systemic omission of the details of who sourced data that were ultimately used by RPOs. This incomplete reporting in the case studies is a problem because it distorts the picture of enabling factors that support the impact work of research performing organisations.

For instance, “the La Trobe University led Comparing Standard Maternity Care with One-to One Midwifery Support (COSMOS) randomised trial […] resulted in a 22% reduction in the proportion of women requiring caesarean section, and a reduction in the proportion of babies requiring admission to a special care nursery”. Here, the data sourcing is attributed to La Trobe University, as researchers from this university presumably collected and managed the data.

An example of non-RPO data sourcing is business data initially collected and held by the government, which were further cleaned and linked up by the Swinburne University industrial economics team. The Business Longitudinal Analytic Database Environment (BLADE) developed from this effort allowed for the generation of insights which influenced industry policymaking. In this example, the government was the entity doing the sourcing of the research data per se, while the RPO came along later in the process to add further value that was then used to develop impact.

In rare occasions, certain stakeholders with which research teams collaborated were noted as the data-sourcing organisation, although the preparation and translation into impacts was attributed to the research teams’ efforts. However, the general picture that emerges from this analysis is that co-creation approaches to conducting research and developing impact from research data are very rare in the Australian case studies we analysed.

## 8.3 Links between data sourcing and field of research

The category of data sourcing, collecting or owning was found to be strongly influenced by the field of research, with more than half (53%) of the variance in data sourcing explained by the field of research in our sample.[85]

### Research Performing Organisations

RPOs were *less* likely to have sourced/collected impact-linked research data in the fields of *Economics*[86], *Information and Computing Sciences*[87], *Interdisciplinary Research*[88] and *Language, Communication and Culture*[89] when compared to other fields of research.

### Non-Governmental Organisations

NGOs where more likely to be the source of impact-linked research data associated with the field of *Language, Communication and Culture*[90] than other FoR categories.

### Quasi-Governmental Organisations

This category, made up of organisations such as established community groups, was more likely to lead the data sourcing/collecting in *Interdisciplinary Research*[91] case studies.

### Government

The government tended to be the owner/collector of impact-linked data related to *Economics* research[92], which was then used by research teams to develop the reported impacts.

### Industry / Business

The fields of *Agricultural and Veterinary Sciences*[93], *Earth Sciences*[94], and *Information and Computing Sciences*[95] were linked to impacts resulting from data collected or owned by industry and business more than other FoR categories.

## 8.4 Links between data sourcing and Go8 membership

Overall, there was a weak association between Go8 membership and the sourcing of the data.[96] We found that 14% or the variability in which type of organisation led the sourcing/collecting of research data could be predicted by Go8 membership. Specifically, RPO-data sourcing was more frequent in EI 2018 case studies from Go8-universities compared to non-Go8 universities in our sample.[97]

[85] $\chi^2(115, N = 715) = 1001435, p < .001, V = .53$
[86] Expected: 96%, observed: 83%, difference: -13%, $p = .034$
[87] Expected: 96%, observed: 33%, difference: -63%, $p < .001$
[88] Expected: 96%, observed: 81%, difference: -15%, $p = .011$
[89] Expected: 96%, observed: 0%, difference: -96%, $p < .001$
[90] Expected: 1%, observed: 100%, difference: +99%, $p < .001$
[91] Expected: 1%, observed: 19%, difference: +18%, $p < .001$
[92] Expected: 1%, observed: 17%, difference: +16%, $p < .001$
[93] Expected: 2%, observed: 11%, difference: +9%, $p < .001$
[94] Expected: 2%, observed: 20%, difference: +18%, $p < .001$
[95] Expected: 2%, observed: 33%, difference: +31% $p = .003$
[96] $\chi^2(5, N = 715) = 13943, p = .016, V = .14$
[97] Expected: 96%, observed: 99%, difference: +3%, $p = .007$

# 9 Comparing Phase I and Phase II findings

In Phase I of this research project, case studies from the UK's Research Excellence Framework (REF) were used to assess research data-linked impact. This section compares the findings from the Phase I findings from UK REF case studies and the Phase II findings from ARC EI 2018 case studies.

## 9.1 Impact type

With one exception, the main types of research data-linked impacts and their proportions were nearly identical between Phase I and Phase II.

*Table 15. Comparison of Impact Type distribution between Australia (Phase II) & UK (Phase I) findings*

| | ***Australia (Phase II)*** | ***UK (Phase I)*** |
|---|---|---|
| **Impact Type** | | |
| Practice Impact | 44% | 45% |
| Other Government / Policy Impact | 16% | 15% |
| Economic Impact | 14% | 13% |
| Public Health Impact | 8% | 5% |
| Other Kind of General Public Impact | 4% | 3% |
| Government Spending / Efficiency Impact | 4% | 6% |
| Environment Impact | 3% | 1% |
| General Public Awareness Impact | 3% | 10% |
| Justice / Crime Reduction / Public Safety Impact | 2% | 2% |
| Other Non-Academic Impact | 1% | 1% |

The exception is that Australian case studies (Phase II) more frequently indicated *Public Health Impacts* (8%) than *Public Awareness Impacts* (3%), while UK case studies (Phase I) had a stronger emphasis on *Public Awareness Impacts* (comprising 10% of all impacts). This stems from a difference in research foci between the Australian and UK samples, with more Australian case studies focused on topics related to medicine/healthcare and natural disasters.

## 9.2 Impact instrument and impact beneficiary

The three most frequent means through which impacts were delivered (i.e., impact instruments) are the same in both the Australian and UK samples. Table 16 shows side-by-side comparisons for the results of the Australian and UK phases of this research, displaying two different possible metrics for *Impact Instruments* and *Beneficiaries*. The metric generally used for this report indicates the percentage of a specific category among all identified instances of the variable, whereas the Phase I report reported primarily using the percentage of cases to which a certain category of the respective variable applied. Both are valid options and are presented here for the sake of completeness.

This written comparison will use the metric focussing on the percentage of categories (left in Table 16) – consistent with the results in the sections above.

*Table 16. Comparison of Australia (Phase II) and UK (Phase I) results, including the percentages of categories among all instances of a variable (allowing for multiple per impact case), and the percentage of cases to which a category of a variable applied.*

| | Percentage of instances of category (multiple per case possible) | | Percentage of cases featuring the category | |
|---|---|---|---|---|
| | **AU** | **UK** | **AU** | **UK** |
| **Impact Instrument** | | | | |
| Improved Institutional Processes / Methods | 33% | 28% | 42% | 40% |
| Report or Static Information | 25% | 23% | 32% | 32% |
| Analytic Software or Methods | 12% | 18% | 15% | 26% |
| Sharing of Tech / Software | 9% | 10% | 12% | 14% |
| Unclear / Uncertain | 8% | 6% | 10% | 8% |
| Other Impact Instrument | 7% | 2% | 9% | 3% |
| Sharing of Raw Data | 4% | 6% | 4% | 9% |
| Searchable Database | 2% | 7% | 3% | 10% |
| Mobile App | 1% | 0% | 1% | 1% |
| **Beneficiary** | | | | |
| Government, Policy, or Policymakers | 28% | 22% | 36% | 42% |
| Industry / Business | 21% | 20% | 28% | 38% |
| Specific Public | 16% | 13% | 20% | 24% |
| Other Organisation | 13% | 6% | 16% | 10% |
| General Public | 8% | 8% | 10% | 15% |
| Professionals | 7% | 26% | 9% | 50% |
| Natural Environment | 5% | 3% | 7% | 6% |
| Unclear / Uncertain | 2% | 0% | 2% | 0% |
| Media | 1% | 1% | 2% | 2% |

For Impact Instruments, the results were very similar between the Australia and UK findings. However, the UK cases were somewhat more likely to feature analytic software and methods as a way to develop impact than the Australian cases we analysed. Moreover, the category *Searchable Database* ranked 8$^{th}$ in frequency in Phase II (2%), while in Phase I it was the 5$^{th}$ most frequent category (7%).

The distribution of Impact Beneficiaries is also very similar between the Australian and UK samples overall, with two major exceptions. First, Australian (Phase II) research results showed almost double the percentage of *Other Organisations* (13% vs 7%). Second, *Professionals* were more than 3.5 times as likely to be beneficiaries in Phase I than in Phase II (26% vs 7%).

Both of these impact beneficiary patterns could perhaps be explained by a difference in processes and areas of focus leading to who benefits from research data. For instance, medical research might lead to impacts benefitting either medical staff by improving their skills (more common in the UK cases), or it might benefit hospitals/clinics in the way they operate (more common in the Australian cases). For Australian researchers and research impact staff, these findings may indicate an under-developed opportunity to deploy research data to enhance the capacities of individual professionals.

## 9.3 IMPACT COUNTER-FACTUAL

Both the Phase I and Phase II findings indicated that remarkably high percentages of reported impacts could not have been generated without research data. However, the UK sample (97%) showed a larger proportion of such cases than the Australian data (93%).

# 10 CONCLUSION

In general, as we found in the Phase I research looking at UK impact cases, impact is rarely delivered directly through research data. Instead, it is developed through different means of extracting value from research data and getting that value into the hands of people and organisations that can use it. This Phase II research highlights this point: *Improved Institutional Processes / Methods* (33%), *Reports or static Information* (25%) and *Analytic Software or Methods* (12%) were found to be the most frequently employed ways of developing impact from research data in both the Australian and UK samples of impact cases. This shows that analysis, curation, product development or other strong interventions are needed to leverage value from research data. These interventions help to bridge the gap between research data and potential users or beneficiaries.

In both Phase I and Phase II, beneficiaries of research data-linked impact are most often *Government, Policy, or Policymakers* (22% and 28%), *Industry /Business* (20% and 21%) and *Specific Publics* (13% and 16%). Government and industry impacts are, in turn, potentially developed into onward insights, services, products, and policies that can lead to broader public impacts. Research data are therefore playing an upstream role within the research and innovation system and linked systems aimed at social, economic and environmental development. This notion is strongly supported by the overwhelming majority of impacts identified in the Australian cases being only indirectly related to the research data (99%). While good data management, open data and streamlined access to data are necessary, further interventions are needed to maximise impact from research data. Extending the use of research data beyond academia requires not only traditional academic research skills, but also capabilities in public communication, entrepreneurship and boundary-crossing. This means that research data usually needs to be processed, communicated or used for developing technologies so that various stakeholders and publics can benefit from them.

The origins of the impact-linked research data from EI 2018 cases were predominantly *Specific Publics* (21%), *Other Organisations* (17%), *Industry / Business* (17%) and the *Natural Environment* (17%). There are also observable links between the data provenance and matching beneficiaries, indicating that certain research tends to benefit the entities from which the data were sourced.

Moreover, for an overwhelming majority of identified research data-linked impacts, *Research Performing Organisations* such as universities seemed to be responsible for data sourcing. In a minority of cases, the collaborative nature of some research projects meant that businesses and other entities remained the owners of the research data involved in the impact case.

In the present study, the *Field of Research* associated with the engagement and impact cases was confirmed to be a statistically significant predictor of outcomes such as impact type. This indicates that there are differences across academic disciplines in impact priorities, approaches and beneficiaries. These disciplinary differences could be fruitful to explore in order to tailor support structures for impact to the needs and potential impacts of particular disciplines and fields of research.

The Phase I and II analyses revealed that research data were a basis for impact in a surprisingly high proportion of impact instances (97% UK and 93% Australia). This suggests that research data may play an essential role in developing impacts that deliver value to society, including in terms of economic value, justice, health, the environment, and other types of impact. If this finding holds true even to a minimal extent, the volume and importance of impacts that could only have been developed through research data is staggering. Therefore, it is essential that opportunities for developing impact from research data are seized and supported.

# 11 References

Jensen, E. & Laurie, C. (2016). *Doing real research: A practical guide to social research*. SAGE: London.

Jensen, Eric A., & Reed, Mark. (2019). *Investigating the Link Between Research Data and Impact*. Zenodo. http://doi.org/10.5281/zenodo.3543505

# 12 ACKNOWLEDGEMENTS

The Institute for Methods Innovation (IMI) would like to express deep appreciation for the time, consideration, and guidance of the ARDC Advisory Group (Dr Kumara Henadeera, Dr Julia Tresidder, Dr Steven McEachern, and Professor Mark Reed) and especially Dr Paul Wong. In addition, we are grateful for the ARDC's, ARC's, CSIRO's and other stakeholders' interest in developing non-academic impacts from research data. The project also benefited from the contributions of the ARC Engagement and Impact case study authors and the open publication of these case studies. We would like to thank the project advisor, Professor Mark Reed, for his guidance in Phase I and support in Phase II of this research. Finally, the IMI researchers contributing to this project include: Dr Eric A. Jensen, Axel Pfleger, Lars Lorenz, Sarah Noles and Dr Aaron Jensen.

# 13 APPENDIX A: METHODS

There were several detailed steps employed to ensure methodological quality in this research project. These steps are detailed in this section of the report.

## 13.1 RESEARCH DESIGN

The planned research evaluates the accounts contained in impact cases for the ARC Engagement and Impact (EI) Assessment 2018. The planned research focuses on the content of ARC EI 2018 impact narratives, investigating how research data delivered positive societal outcomes and what factors enabled such outcomes to develop. The detailed codebook and coder guide are provided in the Appendices of this document.

On the date of collection (04/07/2019) from the ARC database, the total number of impact cases of 279 contained some cases which contained more detailed information about the respective impacts, and others which neither had an ID nor impact information (marked Request Not to Assess [RNTA] = yes). After having extracted the 246 cases containing information, screening for descriptions including key words such as "data" and "dataset" eliminated 57% of all cases, leaving 105 impact cases. The relevance screening furthermore revealed multiple impacts within single cases: a total of 715 individual impact (sub)cases were ultimately identified.

## 13.2 DESCRIPTIVE STATISTICS

In Phase I of this research project, a metric was used for some of the outcome variables which described the frequencies and percentages of cases to which a certain category applied. In this Phase II report, we preferred an analysis which focussed on the frequencies and percentages of respective variables' categories. This metric indicates how often a specific category occurred among all identified items of the respective variable, whereas the Phase I report showed how often a certain category applied to an impact case.

In the comparison between Phases I and II, both lines of results using both metrics were pitted against each other to ensure transparency and comparability.

## 13.3 ASSOCIATION ANALYSES

For the additional analysis related to the Field of Research and the Go8-status as independent categorical variables (i.e. variables containing unrankable text instead of numbers), chi-square tests for association and independence were conducted to determine whether either of the afore mentioned variables could influence all other outcome variables (e.g. Impact Category, Impact Instrument, etc.).

Additionally, Cramer's $V$ was tested to determine the precise nature of the differences and associations if the chi-square test resulted to be statistically significant at a level of $\alpha = .05$. Cramer's $V$ is a measure of association, meaning that it indicates how strongly two variables are associated.[98].

[98] $V = 1$: weak association; $V = 3$: moderate association; $V = 5$: strong association

Measures of association or effect size are vital for the quality of research and data analyses as statistical significance alone can easily be achieved through large sample sizes. Effect sizes provide information on how important findings are, instead of only stating that there are statistically significant findings.

Post-hoc Bonferroni tests helped furthermore identify which categories among the respective independent and dependent variables significantly diverged from the expected values, i.e. which categories were associated with each other. The Bonferroni test takes the z-score resulting from the difference in observed and expected values, and corrects the significance level for determining said association.

It must be added that the Bonferroni correction increases the risk of type II error (falsely accepting the null hypothesis).

## 13.4 INTERCODER RELIABILITY ANALYSIS

The concept of *intercoder reliability* refers to the extent to which independent analysts (or 'coders') evaluating the same content characteristics have reached the same conclusion.

A high level of agreement is taken as evidence that the content analysis has identified characteristics that are objectively evident in the texts being analysed.

The first step in evaluating inter-coder reliability is to have the members of the coding team independently code the same (randomly selected) sub-set of sample cases with overlap between the two (or more) analysts of 10% as the typical benchmark (e.g. see Jensen & Laurie, 2016). In accordance with this standard practice, 10% of the cases analysed for the present research were randomly selected and tested for inter-coder reliability using the statistical test called Krippendorff's Alpha[99] (or 'Kalpha'). 112 units were analysed for this quality assurance step. Variables showing '1' in the Kalpha table indicate perfect agreement between the two analysts.

*Table 17. Krippendorff's Alpha results showing high inter-coder reliability*

| Variable Name | Kalpha |
|---|---|
| Impact Category | 0.85 |
| Searchable Database | 1 |
| Report or Static Information | 0.95 |
| Mobile App | 1 |
| Analytic Software or Methods | 0.87 |
| Improved Institutional Processes / Methods | 0.87 |
| Sharing of Raw Data | 1 |

[99] Krippendorff's alpha is generally considered the most appropriate statistical option for measuring inter-coder reliability for content analyses because it is not affected by missing data, is not sensitive to differences in the number of categories, sample sizes or analysts, and because it can be used regardless of variable type.

| | |
|---|---|
| Sharing of Tech / Software | 0.85 |
| Other Impact Instrument | 0.92 |
| Unclear / Uncertain (Impact Instrument) | 0.64 |
| Impact Pathway | 1 |
| General Public | 1 |
| Specific Public | 0.83 |
| Media | 1 |
| Professionals | 0.74 |
| Government, Policy, or Policymakers | 0.98 |
| Industry / Business | 0.95 |
| Other Organization | 1 |
| Natural Environment | 0.85 |
| Unclear / Uncertain (Beneficiary) | 1 |
| Impact Counter-Factual | 0.94 |
| Data Provenance | 0.83 |
| Data Sourcing | * |

The results show that there were generally very good inter-coder reliability scores across the variables (with one exception, all above .8 Kalpha, which is the established benchmark for good reliability). The one response category with a Kalpha score under .8 ('Unclear / Uncertain (Impact Instrument') is a low priority variable, which is not used in the primary analysis.

For the final data analysis, the differences between both coders were resolved through discussion, so that the impact cases used to determine the inter-coder reliability could be integrated into the general dataset for the main statistical analysis.

*Krippendorff's alpha indicates the likelihood of coder agreement not having occurred by chance, where 1 means there is no chance coder agreement happened by chance, and 0 means coder agreement occurred completely by chance. Although the variable *Data Source* had eight categories, only one was actually used in the coding process. For this variable, the algorithm assumes the existence of only one category, which would by this logic be the only possible one to code for. Following this assumption, coder agreement could have happened completely by chance. In reality, however, the coders had perfect agreement when choosing from eight different possible categories.

# 14 APPENDIX B: CODING GUIDE

The following are detailed instructions for coding each Impact Case, using the spreadsheet you have been provided. Before beginning the coding process, please read this document carefully from top to bottom. You should also use this information as reference during coding.

Please note that most categories have an “other” option. If at any time you are unsure of how to code an element of an article or believe that you have found something unaccounted for within the current coding scheme, please use the “other” code.

## General Notes

- Any impacts generated through mechanisms that do not involve research data are out of scope for this analysis and should be excluded.
- There may be more than one impact represented in a case study. You will consider all of the analytic steps outlined below for each impact identified.
- Err on the side of inclusion / positive identification of a category. (That is, be on the liberal side in terms of allocating content to a category if you feel a bit uncertain).

## Pre-Entered Coding Information

On the “Overview” sheet of the coding Excel file you see a list of all impact cases you will need to code. Every sheet of the Excel file is its own impact case, uniquely identifiable by an ID number. On each impact case sheet you will find some general information about the case at the top and the information that needs to be coded below. Each case is divided by paragraphs and each paragraph needs to be coded separately, using the coding approach outlined below. Your first task will be to identify whether a paragraph contains relevant information, in which case it must be coded further, or not, in which case you should skip to the next paragraph.

**Please do not edit any fields of the coding sheet, except for the coding area! Any other changes to the coding sheet will greatly impact the efficiency of analysis.**

## Definition of Impact

For the purposes of this study, Impact is defined as any positive effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia.

# 1 Screening for relevance

The first step when coding a case study is to determine whether its paragraphs meet the criterion for inclusion in the analysis. In some case studies, you may find that there is actually no 'research data' present. For example, the 'data' referenced may actually have nothing to do with research. Or the impact that is described may be unrelated the to the research data. In such cases (when there is no research data-linked impact) you should code a negative response (0) in the "valid paragraph" column and leave all other columns for that paragraph (row in the coding sheet) blank.

## 1.1 Codes

0 = No, the paragraph is not valid. There are no impacts/benefits generated via research data mentioned in this paragraph.

1 = Yes, the paragraph is valid. There is at least one impact/benefit generated via research data mentioned in this paragraph.

## 1.2 Text Cleaning

Since the coding sheets have been automatically generated, you will occasionally find either fragmented paragraphs or rows only containing headings or references.

For fragmented paragraphs, merge the text into the top row of the fragmented paragraph and delete the fragmented rows afterwards. For headings, references or other short non-paragraph text, also delete the entire row. Please always make sure to **delete** the entire row in Excel and do not modify the UID in any way.

## 1.3 Relevance Screening

If a paragraph has been identified as valid, you will need to screen it for relevant text segments. Copy all segments that directly deal with the research topic from the "Original Paragraph" column to the "Edited Paragraph" column. This step can be skipped for all paragraphs that are **NOT** marked as valid.

Use the 3rd column (Edited Paragraph) to edit the paragraph and do not modify the 2nd column (Original paragraph).

### 1.3.1 Multiple Impacts per Paragraph

Should a paragraph contain more than one impact, you may split it into multiple paragraphs. This should only be done if you are absolutely sure that it is required for proper coding.

To split a paragraph, insert a new row in Excel below the paragraph you are splitting. Copy the UID from the original paragraph and add "_1", "_2", "_3" and so forth for every additional paragraph to the end of the UID (example: 672_5_C1 becomes 672_5_C1_1). Leave the "Original Paragraph" column unchanged for the original paragraph and leave it blank for each split paragraph. Use the "Edited Paragraph" column as usual, splitting your paragraph accordingly across rows. Use the original paragraph's "Edited Paragraph" column as your first split paragraph.

### 1.4 The Same Impact Repeating

Occasionally you will come across paragraphs that are essentially an abstract or summary of the entire case, leading to an impact being repeatedly mentioned. Proceed as usual for these cases, but if you are absolutely sure the impact is repeated, add a comment in the “comments” column, referencing the paragraph that is being repeated as such: REPEATS [paragraph ID]

# 2 Reviewer Agreement

After step 1, you will switch coding sheets with the other coder(s) and check their work on validating paragraphs, relevance screening and text cleaning. If you are in disagreement with their work, please code a 1, otherwise code a 0. Note that disagreement has to be substantial to the research, i.e. do not indicate disagreement for minor wording issues or typos.

0 = I disagree with the other coders preparatory work on this paragraph.

1 = I agree with the other coder.

# 3 IMPACT CATEGORY

Once an impact has been identified, the first step is to code for the type of impact, or **impact category**. Each impact should be assigned one, and only one, impact category. (Note: If there is more than one impact category, code the most relevant one and check whether the other impacts are specifically discussed more clearly in other paragraphs of the current case study. IF NOT, then add an additional row to code the second type of impact. If YES, then code this other impact type in the more detailed paragraph).

## 3.1 DEFINITION OF IMPACT CATEGORY

The type of non-academic benefit / outcome that has been generated via research data.

## 3.2 IMPACT CATEGORY CODES

1 = Government Spending / Efficiency Impact

2 = Other Government / Policy Impact

3 = Practice Impact

4 = General Public Awareness Impact

5 = Justice / Crime Reduction / Public Safety Impact

6 = Public Health Impact

7 = Economic Impact

8 = Environment Impact

9 = Other Kind of General Public Impact

10 = Other Non-Academic Impact

-98 = Unclear / Uncertain

## 3.3 IMPACT CATEGORY CODE DEFINITIONS

### 3.3.1 Government Spending / Efficiency Impact **(1)**

Reducing cost of delivering government services; increasing impact/quality of government service without raising cost.

Example

Research data reveals a way of reducing the cost of having criminals in prison by moving them to house arrest and work at an earlier point in their sentences, which is then implemented by prisons.

### 3.3.2 Other Government / Policy Impact **(2)**

Changing public policy or government regulations, or how either of these are implemented.

Example

Research data reveals the need and policy uptake for more ambitious carbon emission targets within specific sectors, such as transport, if the Government's larger 2050 goal of being carbon-neutral is going to be reached.

### 3.3.3 Practice Impact **(3)**

Changing the ways that professionals operate; changing organizational culture; improving workplace productivity or outcomes; improving the quality of products or services through better methods, technology, understanding of the problems, etc.

Example

Research data used as part of training for schoolteachers, helping them to deliver more effective math teaching for children at Key Stage 2.

### 3.3.4 General Public Awareness Impact **(4)**

Improving public knowledge about a topic or increasing public visibility or attention for an issue.

Example

Research data reveal the public health risk of vaping and an improved understanding within a general public.

Research data provided through a website for members of the public to self-assess on a given variable having an impact on general public awareness.

**Do not include:** Cases that just outline a potential for general public awareness impact, without stating any change or impact actually taking place.

### 3.3.5 Justice / Crime Reduction / Public Safety Impact[100] **(5)**

Reducing crime; Increasing efficiency in reducing crime; Improving justice outcomes (i.e. fairer; less cost; better social outcomes).

Example

Research data highlight a problem with the way that scientific data are communicated during criminal proceedings, which results in the scientific findings being misunderstood by both judges and juries. The research leads to reforms in the communication process.

### 3.3.6 Public Health Impact[101] **(6)**

Improvements to the health of the population or a part of the population.

Example

[100] Justice refers to criminal justice impacts; Public Safety means a program(s) carried out or promoted by a public agency for public purposes involving, directly or indirectly, the protection, safety, law enforcement activities, and criminal justice system of a given political area.

[101] The areas of public health responsibility include (1) assuring an adequate local public health infrastructure, (2) promoting healthy communities and healthy behaviours, (3) preventing the spread of communicable disease, (4) protecting against environmental health hazards, (5) preparing for and responding to emergencies, and (6) assuring health services.

Research data show a pattern of communicable disease transmission that reveals inadequate sanitation in a particular part of a city. Once identified, this problem is addressed, thereby improving population health.

### 3.3.7 Economic Impact **(7)**
Improvements to the economy or overall financial/economic situation.

Example

A particular company uses research data to more effectively target its sales efforts, leading to increased revenue.

### 3.3.8 Environmental Impact **(8)**
Improvements in the natural environment, or reductions in threats or harm.

Example

Research data reveal new, more effective ways to remove pollution from rivers, which are then applied through a project.

### 3.3.9 Other Kind of General Public Impact **(9)**
Benefits for the general public (not professionals/government) that are not explicitly state above.

### 3.3.10 Other Non-Academic Impact **(10)**
REF eligible non-academic impacts not following into any of the categories above. That is, cannot include academic publications or improvements to the teaching within a researcher’s own institution

### 3.3.11 Unclear / Uncertain **(-98)**
Not enough detail or clarity to clearly identify.

# 4 Impact Instruments

After having identified the nature of the impact, the next question to address is the particular mechanism by which the data created an impact. Multiple impact instruments can be identified for each impact, that is, “impact instruments” is multi-categorical. Each type of impact instrument is split into a separate variable, e.g., II_A, II_B, etc. Please code for the presence or absence of each instrument in generating the impact.

## 4.1 Definition of Impact Instruments

How research data were used to generate impact, that is, the nature of the intervention, the means, or the impact generating activity.

## 4.2 Impact Instruments Sub-Category Codes

0 = No, the impact was NOT generated with this type of instrument.

1 = Yes, the impact was generated with this type of instrument.

## 4.3 Impact Instrument Sub-Categories

### 4.3.1 Searchable Database

A database that can be accessed to view the research data in a dynamic way (that is, offers ability to select variables/filters, allowing for customised information to be accessed by users to use for their own purposes).

Examples

- Research data placed on a website to allow users to search for information relevant to their location.
- **EXCLUDE:** Pre-prepared analyses that show the conclusions or implications in a format that is ready for the end user to employ without further effort.

### 4.3.2 Report or Static Information

Report containing pre-analysed/curated information, a static database, results tables or other methods of presenting the research data as processed information to be used without customisation or filtering of the data.

Examples

- Report presenting analyses of the research data.
- Research data converted into infographics.
- Data tables published in a report.
- Media coverage or media interview.
- **EXCLUDE:** Mechanisms allowing users to filter the data or results or to search through it looking for their own insights or to conduct their own data analysis.

### 4.3.3 Mobile App

An application designed for smartphone or tablet to access the research data or an analysis/results of the data.

Example

- Smartphone app displaying the distribution of certain crimes on a dynamic map of a local area based on research data.

### 4.3.4 Analytic Software or Methods

Research data used to generate or refine software or research/analytic methods or statistical models.

Examples

- Software capable of detecting anomalous tissue samples more effectively as part of cancer screenings.
- Cloud software to do automatic analysis of social media content to identify potential terrorist threats.

### 4.3.5 Improved Institutional Processes / Methods

Research data used to make an institution's way of operating better/more efficient or more effective at delivering outcomes.

Example

- Research data identified faults in recruitment and selection processes resulting in under-recruitment of women to certain kinds of jobs. The research data shows the way to a better process with lower potential for gender bias.

### 4.3.6 Sharing of Raw Data

Research data has an impact via being shared with others (in raw or minimally anonymised form) outside of the research team that generated the data so that they can do something with it (e.g. further analysis, etc.).

Example

- A research data set was prepared for publishing as open data and shared on a national repository. Another researcher accessed this data and conducted an analysis, leading to new insights that delivered positive impact.

### 4.3.7 Sharing of Tech / Software

The research data have an impact via sharing technology or software that was created using the research data or that uses the research data somehow.

Example

- Research data used to refine a text analysis software tool, which is developed within an open source framework and published on GitHub. Another developer adapts this tool to deliver an automated text analysis service that makes city government customer service more responsive to public comments.

### 4.3.8 Other Impact Instrument

A clearly identifiable impact instrument that does not fit into any of the categories listed above.

### 4.3.9 Unclear / Uncertain

Impact instrument that is not detailed enough to clearly place into any pre-specified category.

## 4.4 Coding Note

We are coding for the presence or absence of a coding category. It’s possible that an impact instrument sub-category was used more than once for a given impact. This will be coded no differently than if the impact instrument was only used once.

# 5 Impact Pathway

This code is focused on whether direct experience or visibility of the research data was needed for the impact to be generated, that is, whether the research data was used to create something beneficial, or if creation / dissemination of the research data was the benefit.

## 5.1 Definition of Impact Pathway

The role of data in generating the impact, i.e., whether the impact was generated 'directly' though engagement with the research data or 'indirectly' through creation of some other product/service.

## 5.2 Impact Pathway Codes

1 = Indirect Impact Pathway

2 = Direct Impact Pathway

-98 = Unclear / Uncertain

## 5.3 Impact Pathway Code Definitions

### 5.3.1 Indirect Impact Pathway **(1)**

Research data used to create something that has impact (data as input to impact generating activity). In this category, the beneficiary's contact with the research data is mediated through some other mechanism, service, product, report or presentation.

Example

- Research data used to inform an investigative journalism story, which reveals important changes needed in the way that eligibility for disability benefit is evaluated by the government. (End users only encounter the journalistic story, not the data directly)

### 5.3.2 Direct Impact Pathway **(2)**

Research data per se used as the impact generating intervention (e.g. researchers use data as output). In this category, the beneficiary has direct contact with the research data.

Example

- Genetic research data published online in a searchable format, which people can access and use to identify their risks for certain diseases. (End users are able to access the data more or less directly, without filtering through e.g. a health news website that simplifies the results)

### 5.3.3 Unclear / Uncertain **(-98)**

Not enough detail or clarity to clearly identify pathway type.

# 6 Beneficiary

This code refers to the nature of the people or organizations that benefited from the research data. As with impact instruments, more than one beneficiary can be identified for each impact, that is, "beneficiary" is multi-categorical. Each type of beneficiary is split into a separate variable, e.g., BEN_A, BEN_B, etc. Please code for the presence or absence of each type of beneficiary.

## 6.1 Definition of Beneficiary

The type and nature of the people, organizations, etc. that benefited from the research data, directly or indirectly.

## 6.2 Beneficiary Sub-Category Codes

0 = No, the impact did NOT have this type of beneficiary.

1 = Yes, the impact had this type of beneficiary.

## 6.3 Beneficiary Sub-Categories

### 6.3.1 General Public

Unspecified public as beneficiary; or society-wide, community-wide or a national or regional audience.

- **Includes:** benefits for the well-being of a city, town, or neighbourhood.
- **Excludes:** Benefits to a business, government, non-governmental organization, etc.

### 6.3.2 Specific Public

Particular demographic category of non-professional/non-governmental/non-business beneficiary specified (e.g. children/mothers/art museum visitors/etc.)

Examples

- Stakeholders who will be directly affected by a new government policy or business development.
- Children.
- Mothers.
- Art museum visitors.
- **EXCLUDE:** If the stakeholders are taking an interest or is affected due to their professional role / job.

### 6.3.3 Media

Research data improving media services, enabling data journalism, resulting in news coverage or news insights, documentaries or entertainment media.

Examples

- Improved (e.g. more accurate, detailed) news coverage about a topic relating to the research data.
- Offering media consumers new insights or access to information.
- Used to inform storyline for a TV documentary.

### 6.3.4 Professionals

Improved capacities, skills, employment options, increased salaries or benefits, greater influence, etc.

Examples

- Research data used in the training of factory workers done in a different way to reduce error rates.
- Research data used to help people training to be computer coders to develop their skills more quickly.
- **NOTE:** This category focuses on the level of people/individuals gaining improved capacities (not organisations or companies as a wider entity).
- **INCLUDE:** If these improved capacities would go with the people if they switch jobs.
- **EXCLUDE:** If the improved capacity belongs to the company/organisation and stays with the company/ organisation when the professional leaves.

### 6.3.5 Government, Policy, or Policymakers

All levels of government (e.g. affecting how government delivers services, prioritises etc) or government policy (e.g. work cited or underpinning legislation, regulation or new policy initiatives) or policymaking (e.g. expert committees or feeding into city or government department research designed to inform policy or regulations)

Examples

- Research data used to organise the schedule of trash pickups more efficiently, thereby saving time/resources.
- A government policy about reducing household energy use is designed based in part on the research data.
- Research data are used as part of the evaluation of existing government programmes or services to highlight where they need to be improved.

### 6.3.6 Industry / Business

Improving business outcomes, profits, service/product quality, reducing environmental impact, etc.

Examples

- Increased revenues for a particular company
- Expansion of a sector within an industry, a business or a set of businesses
- Research data used by a market research company to improve its proprietary methods of segmented communication to consumers.
- Research data used to test and refine proprietary software designed to more effectively target cancer treatments.

### 6.3.7 Other Organization

Non-governmental / Non-business organization not covered by the above categories.

Examples

- Private non-profit hospital.
- Non-governmental advocacy organisation.
- Registered charity.

### 6.3.8 Natural Environment

Improved environmental outcomes, or reduction in damage/threat.

Examples

- More efficient use of water.
- Lower carbon footprint.

### 6.3.9 Unclear / Uncertain

Not enough detail provided to clearly identify the nature of the beneficiary.

## 6.4 Coding Note

We are coding for the presence or absence of a coding category. It is possible that a there was more than one beneficiary within the same sub-category. This will be coded no differently than if there were only one beneficiary within that sub-category.

# 7 IMPACT COUNTER-FACTUAL

This code assesses whether, based on your judgment and the information available within the specific impact case study paragraph, you feel the impact could have or would have occurred without the data.

## 7.1 IMPACT COUNTER-FACTUAL DEFINITION

Coders answer to the question: *Based on the information in the case study, was the research data required for the identified impact to exist*?

## 7.2 IMPACT COUNTER-FACTUAL CODES

0 = No

1 = Yes

2 = Partial No

-98 = Unclear / Uncertain

## 7.3 IMPACT COUNTER-FACTUAL CODE DEFINITIONS

### 7.3.1 No **(0)**

The research data were not essential for the identified impact to develop.

### 7.3.2 Yes **(1)**

The research data were the only pathway to the identified impact.

### 7.3.3 Partial No **(2)**

The research data were essential for some of the identified impact or for some of the beneficiaries, but not all.

### 7.3.4 Unclear / Uncertain **(-98)**

Not enough detail or clarity to clearly whether or not the impact could have occurred without the data.

# 8 Data Provenance

This code assesses where, based on your judgment and the information available within the specific impact case study paragraph, the data provenance is. Only a single provenance can be coded per impact. Should the data supporting the impact have multiple provenances, please code the most relevant one.

## 8.1 Definition

Origin of the data underlying the identified impact. This refers to the data provenance, that is, where or from whom the data were sourced or collected. *Data provenance does NOT refer to the owner, author, data collector or distributor of the data* (this is addressed in the next section on Data Sourcing).

## 8.2 Data Provenance Codes

1 = General Public

2 = Specific Public

3 = Media

4 = Professionals

5 = Government, Policy, or Policymakers

6 = Industry / Business

7 = Other Organization

8 = Natural Environment

-98 = Unclear / Uncertain

## 8.3 Data Provenance Code Definitions

### 8.3.1 General Public **(1)**

Unspecified public/demographic category as data provenance; or society-wide, community-wide or a national/regional audience.

Examples

- Broadly sourced public opinion or community wide statistics.

### 8.3.2 Specific Public **(2)**

Specified public/demographic category of e.g. customers/children/mothers/art museum visitors/ etc.

Examples

- Data generated from people shopping online, shopping behaviour or media statistics about consumption or transactions.
- Art museum visitors (e.g. children/mothers) for educational or entertainment purposes.
- **EXCLUDE:** If the stakeholders are taking an interest or are affected due to their professional role / job or being involved in governmental, industry/business affairs.

### 8.3.3 Media **(3)**

Media services, (data) journalism, press releases, news coverage or news insights, documentaries or entertainment media as data provenance.

Examples

- Data on news coverage, (social) media or other news / entertainment information.
- **EXCLUDE:** Behaviour on media consumption. The main research object (i.e. news consumption in teenagers) acts as data provenance.

### 8.3.4 Professionals **(4)**

Professionals with specific capacities, skills, employment options, increased salaries or benefits, greater influence, etc. as data provenance.

Examples

- Factory workers or people training to be computer coders.
- **NOTE:** This category focuses on the level of people/individuals as data provenance, not organizations or companies as a wider entity.
- **INCLUDE:** If the professional capacity would go with the people if they switch jobs.
- **EXCLUDE:** If the professional capacity belongs to the company/organization and stays with the company/ organization when the professional leaves.

### 8.3.5 Government, Policy, or Policymakers **(5)**

All levels of government (e.g. affecting how government delivers services, prioritizes etc.) or government policy (e.g. work cited or underpinning legislation, regulation or new policy initiatives) or policymaking (e.g. expert committees or feeding into city or government department research designed to inform policy or regulations).

Examples

- Government controlled actions, such as trash pickup.
- Data on politics, policies, policymakers, government programmes or services.

### 8.3.6 Industry / Business **(6)**

Businesses/industry, business outcomes, profits, services/products or similar as data provenance.

Examples

- Particular company or sector within an industry, a business or a set of businesses.

#### 8.3.7 Other Organization **(7)**
Non-governmental / Non-business organization not covered by the above categories.

Examples

- Private non-profit hospital.
- Non-governmental advocacy organization.
- Registered charity.

#### 8.3.8 Natural Environment **(8)**
Environmental / natural data provenance.

Examples

- Climate, soil or other data directly gathered from natural sources.
- **Note:** Data is most likely collected by researchers or industry, however keep in mind where the data provenance sits. For example, if a researcher collects soil data, the data provenance is natural / environmental.

#### 8.3.9 Unclear / Uncertain **(9)**
Not enough detail provided to clearly identify the data provenance.

# 9 Data Sourcing

This code assesses who, based on the information available within the specific impact case study paragraph, sourced or collected the data supporting the impact. Only a single data source can be coded per impact. Should the data supporting the impact have multiple sources, please code the most relevant one.

Note: If you are unclear about an organisation's/institution's status, please briefly look them up.

## 9.1 Definition

Source of the data underlying the identified impact. This refers to who collected, curated, sourced or owned the data used in the data focused research impact, i.e. which entity is responsible for taking the data from the data provenance. Data source does NOT refer to the provenance of the data.

## 9.2 Data Source Codes

1 = Research Performing Organization (RPO)

2 = Research Funding Organization (RFO)

3 = Non-Governmental Organization (NGO)

4 = Quasi-Governmental Organization

5 = Government

6 = Industry / Business

7 = Other data source

-98 = Unclear / Uncertain

## 9.3 Data Source Code Definitions

### 9.3.1 Research Performing Organization (RPO) **(1)**

Organization performing research or housing research activities, such as universities, institutes, academies, or similar.

Examples

- University researchers collect data on public attitudes towards a topic.

### 9.3.2 Research Funding Organization (RFO) **(2)**

Any organization funding research and sourcing or collecting data.

Examples

- Funding organization collecting data on funded research projects, which is used in a data driven research impact case.

### 9.3.3 Non-Governmental Organization (NGO) **(3)**
Non-Governmental Organizations, such as environmental or health organizations, associations or other citizen founded organizations.

Examples

- Amnesty International collects and shares a dataset on human right violations.

### 9.3.4 Quasi-Governmental Organization **(4)**
A corporation or body with a public mandate that is directly supported by the government. Also known as 'arms-length' bodies that operate independently from the government. Quasi-Governmental Organizations are usually involved in providing oversight, funding, or have accountability for public benefit.

Examples

- Arts Council for the Arts, Australia, collects data on museum visits.

### 9.3.5 Government **(5)**
All levels of government, public services, or policymaking (e.g. expert committees, government departments or other governmental entities).

Examples

- The state traffic department collects and shares a dataset on road usage or traffic.

### 9.3.6 Industry / Business **(6)**
Businesses / industry, small companies, or for-profit organizations providing services or products.

Examples

- A large supermarket chain collects and shares a dataset on consumers shopping behaviour.

### 9.3.7 Other data sources **(7)**
Any other data sources, which can be specifically identified, but are not covered by any of the above options.

### 9.3.8 Unclear / Uncertain **(-98)**
Not enough detail provided to clearly identify the data source.