Investigating the Link Between Research Data and Impact

Lead Author:

Dr Eric A. Jensen

eric@methodsinnovation.org

Lead Expert Advisor:

Professor Mark Reed

Fast Track Impact

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, drawing on existing impact case studies from the UK Research Excellence Framework.

Project overview

The research involved analysing impact cases from the UK's Research Excellence Framework (REF). These cases were sifted to only review high scoring cases with a strong emphasis on 'data'. Relevant text to this research was extracted from the larger impact narratives. A content analysis was conducted to identify patterns, linking research data and impact in the narratives. This analysis achieved a high level of reliability, based on established methodological standards.

148

Most explicitly data-focused REF2014 impact cases were identified.

20%

Of the sample was used to develop an analytic framework.

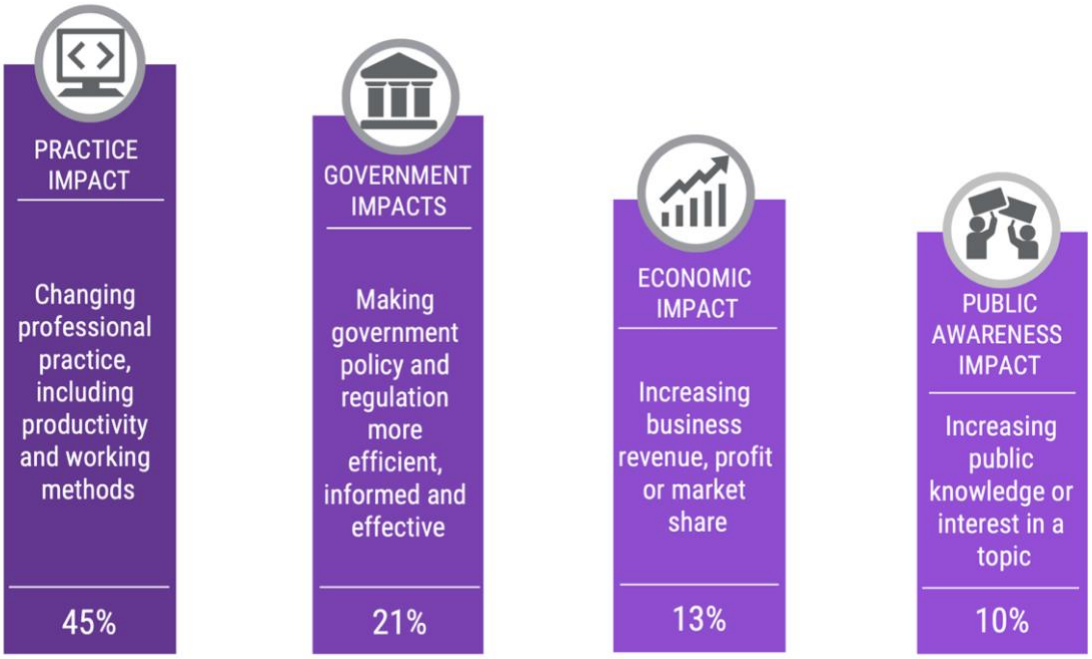
What type of impact was developed from research data?

The most prevalent type of research data-driven impact related to *Practice* (45%). This category of impact includes changing the ways that professionals operate, changing organizational culture and improving workplace productivity or outcomes. It also includes improving the quality of products or services through better methods, technology, understanding of the problems, etc.

Government impacts were the next most prevalent category identified in this research (21%). This category includes reducing the cost to deliver government services, enhancing the effectiveness or efficiency of government services and operations and providing input into government planning, decision-making and policymaking.

Other relatively common types of research data-driven impacts were *Economic impact* (13%) and *General Public Awareness* impacts (10%).

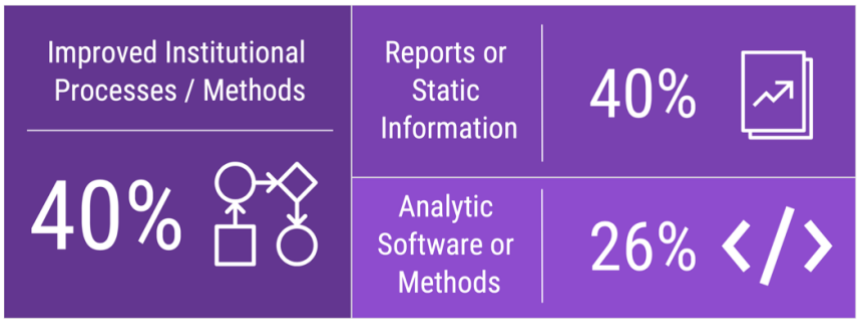
What type of impact was developed?



How was impact developed from research data?

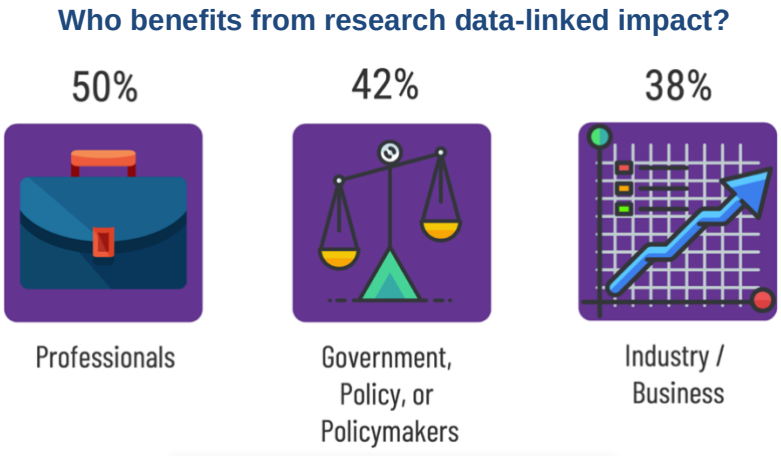
Impact from research data was developed most frequently through *Improved Institutional Processes / Methods* (40%). This relates to making an institution's way of operating better, more efficient or effective at delivering outcomes. The second most common way of developing impact was via a *report* (32%) of some kind, that is, pre-analysed or curated information. *Analytic Software or Methods* (26%) comprised the third most frequently used way of developing impact. Here, research data are used to generate or refine software or research and analytic methods.

Which pathways led to data-linked impact?



Who benefited from the research data-linked impact?

Professionals (50%), Government, Policy, or Policymakers (42%) and Industry / Business (38%) were the most common types of beneficiaries from the research data-linked impact. This finding is partly explained by a two-step flow of research data-linked impact that ultimately reaches publics or wider non-academic stakeholders. While intermediaries such as professionals, policymakers and industry are primary beneficiaries or users of the research data-based impact, they in turn use what they have gained to develop insights, services, products and policies that deliver broader public impacts.



Looking at patterns in this analysis, the following correlations were identified:

a. Searchable databases tended to be used with the general public (r = .22), while 'enhancing institutional processes / methods' is not (r = -.26).
b. Analytic software (r = .23) and 'improved institutional processes / methods' (r = .32) were used more to develop impact with industry / business.
c. Sharing of raw data was more often an impact development pathway with environmental impacts (r = .2) than other types.

Conclusions

The analysis found that research data on their own rarely develop impact, but instead they require analysis, curation, product development or other strong 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.

In the same sense, the impact of research data can be engineered, through closer links between government, industry and researchers, capacity building for researchers to effectively use research data to develop impact and capacity building for potential beneficiaries to establish links with researchers and to access and make sense of useful sources of research data that can be adapted to serve new purposes. Moreover, the way that research data is made available, and the nature of the support available, can affect how feasible it is to use that research data to develop new and creative pathways to impact.

Finally, there were surprisingly high 'uniqueness' scores for the impacts linked to research data (97%), suggesting that most of the research-data linked REF-reported impacts may have only been possible to develop through research data. However, limitations inherent in REF impact case studies have to be taken into account before drawing firm conclusions on this point.

TABLE OF CONTENTS

1 INTRODUCTION

This report highlights key trends and patterns evident in the research conducted for the Australian Research Data Commons (ARDC) about the role of research data in developing impact.

After screening for only high scoring cases, keywords (e.g. 'data' and 'database') were used to search for full impact narratives with high relevance to the present research within the UK's wider Research Excellence Framework (REF) impact case database. The 148 most explicitly data-focused impact cases were identified through this search.

For the analysis, a subset of the sample (approx. 20%) 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.

The categories were applied systematically through a well-established social research method known as 'content analysis'. This yielded quantifiable results, underpinned by intercoder reliability checks on randomly selected subsets. To keep the content analysis focused, impact-related text passages were extracted manually from the longer impact narratives before coding. Finally, the different dimensions uncovered through the quantitative analysis were supplemented by the extraction of case examples.

148

Most explicitly data-focused REF2014 impact cases were identified.

20%

Of the sample was used to develop an analytic framework.

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 the nature of the impact pathways associated with research data is defined in both quantitative and qualitative terms. The final results section reveals the findings relating to whether research data offers a unique pathway to impact. These findings are placed in context with conclusions and recommendations.

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. 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 (in line with the REF definition of impact).

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.

2.1 TYPES OF IMPACT IDENTIFIED

We begin by looking at the different types of impact that were identified in an analysis of REF impact cases that were associated with research data. The initial analysis of the cases revealed a set of different types of impact, defined in Table 1.

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	Greater revenue, profit or market share is developed for a company or sector using research data.
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.
Other Non-Academic Impact	REF-eligible non-academic impacts not falling into any of the categories above. (i.e., cannot include academic publications or improvements to teaching within a researcher's own institution, as these would not be REF-eligible).
Unclear / Uncertain	Not enough detail or clarity to clearly identify.

Table 1. Identified types of impact

2.2 PREVALENCE OF DIFFERENT IMPACT TYPES

The most prevalent[1] types of research data-driven impact in our sample were related to *Practice* (45%) and *Government* (21%), which includes both *Government Spending / Efficiency* (6%) and *Other Government / Policy* impacts (15%).

Likewise, other types of research data-driven impacts such as *Economic* impact (13%) and *General Public Awareness* impact (10%) were also represented in a noteworthy minority of cases.

Impact Type	Percentage
Practice Impact	45%
Government Spending / Efficiency Impact	6%
Other Government / Policy Impact	15%
Economic Impact	13%
General Public Awareness Impact	10%
Public Health Impact	5%
Other Kind of General Public Impact	3%
Other	4%

Table 2. Prevalence of different types of impact associated with research data

Practice impacts

The findings show that 45% 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 the organisational culture; improving workplace productivity or outcomes or improving the quality of products or services through better methods, technology, understanding of the problems.

For example, as a result of the application of research data collected using the Work-Related Quality of Life Scale (WRQoL)[2] scale, there was a marked improvement in workplace wellbeing documented in an impact case study:

> Survey results at organisation level “allowed us to focus on sickness absence", prompting the introduction of a new sickness absence procedure: "days lost […] for stress/mental ill health for academic staff have more than halved from 1453 days to 604 days, a[n annual] saving estimated to be in the region of £100,000”.

Government and policy impacts

21% of impacts were related to one of the following developments:

- Reducing the cost of delivering government services
- Increasing impact/quality of government services, without raising costs

[1] REF2014 impact case content with impact dimensions that fit in more than one field were categorised for each impact separately.
[2] Work-Related Quality of Life Scale

- Changing public policy or government regulations

An example of a case that focuses on reducing the cost of delivering government services or increasing impact/quality of government services without raising costs involved the UK government agency, the Office for National Statistics (ONS). Here, an analysis of research data from the Office for National Statistics was used as a basis to develop a 'disclosure framework and software package' for the release of microdata to a vast range of end-users, including central and local government departments, and the health service. That is, ONS data were used to develop methods and software that in turn could be used to leverage value from ONS data for wide range of stakeholders across government.

Economic Impact & General Public Awareness Impact

Research data were also used to develop impacts on both the economy (13%) and general public awareness (10%). Examples of economic impact included cases that delivered health care cost savings while improving patient health outcomes. General public awareness impacts included changes to how people perceived and understood information related to contemporary topics, such as bullying and social media use, among others. Such general public-oriented impacts often focused on improving public knowledge about a topic or increasing public visibility or attention for an important issue.

3 How impact develops from research data

This section focuses on the ways that impact was leveraged from research data.

3.1 How impact was developed

The impact is developed from research data in several different ways. Here, we analyse the nature of these different interventions, which could be understood as the means of developing impact, or the impact generating activities. The impact development approaches we identified are summarised in the table below.

How impact was developed	Description
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.
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.

Table 3. Identified ways of developing impact

3.2 PREVALENCE OF IMPACT DEVELOPMENT PATHWAYS

This section explores the frequency with which the different ways of developing impact appeared in the cases analysed. The most common[3] ways of developing impact from research data were 'Improved Institutional Processes or Methods' (40%), 'Reports or static information' (32%) and 'Analytic Software or Methods' (26%).

How impact was developed	Percentage
Improved Institutional Processes or Methods	40%
Reports (or other static, curated information)	32%
Analytic Software or Methods	26%
Searchable Database	10%
Sharing of Raw Data	9%
Sharing of Tech / Software	14%
Other Impact Instruments	4%

Table 4. Prevalence of Impact Development Pathways

Institutional Processes or Methods

Improving **'**institutional processes or methods' was a major pathway to developing impact. One example of this category impact development comes from research on regulatory practice.

> "[On regulatory practice] Raab's research has also had a more direct impact on the regulation of information privacy, informing specific policies and frameworks in the UK and beyond. [...] His distint contribution [to the Scottish Government's Identity Management and Privacy Principles report] is evident in principles that warn against discrimination and social exclusion in identification processes; that encourage organisations to raise awareness of privacy issues; and that promote transparency [...] These have been widely cited and applied in Scottish Government policy".

Other examples where research data were used to enhance institutional processes include changes in data privacy practices and enhancements in the transparency of companies' operating procedures.

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.

[3] 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.

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 the previous example showing a ‘government’ impact relating to the UK’s Office for National Statistics. This is because the way that the researchers developed impact involved using research data to create a “disclosure framework and software package”, which was then used to develop impact.

> “Analysis of research data from the Office for National Statistics used as a basis to develop a 'disclosure framework and software package’ for the release of microdata to a vast range of end-users, including central and local government departments, and the health service.”

That is, the mechanism for developing impact in this case was improved methods for data disclosure and new software to implement those improved methods.

4 THE BENEFICIARIES OF RESEARCH DATA-LINKED IMPACT

This section focuses on who is benefiting from research data-linked impacts, and how. Here, we investigate the people and organisations that benefited from the research data-linked impacts examined in this study.

4.1 DEFINING RESEARCH DATA-LINKED IMPACT BENEFICIARIES

The table below contains definitions of the people and organizations that benefited from research data-linked impacts, whether directly or indirectly.

Beneficiary	Description
General Public	Unspecified public as beneficiary; or society-wide, community-wide or a national or regional audience.
Specific Public	Particular demographic category of non-professional/non-governmental/non-business beneficiary 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.

Table 5. Identified Impact Beneficiaries

4.2 RESEARCH DATA LINKED IMPACT BENEFICIARIES

This section addresses how common[4] the categories of beneficiaries are within research data-linked impacts we have identified.

Impact beneficiaries	Percentage
Professionals	50%
Government, Policy, or Policymakers	42%
Industry / Business	38%
Specific Public	24%
General Public	15%
Other Organization	10%
Natural Environment	6%
Media	2%

Table 6. Prevalence of different types of beneficiaries in the analysed impact cases

Professionals were the primary beneficiaries in 50% of the identified research data-linked impacts in this study. They benefitted through improved professional capacities and skills, some of which they were able to use, in turn, to deliver broader public impacts. Professionals also benefitted personally from enhanced employment options and opportunities, as well as increased salaries or benefits, and other positive outcomes.

Government, policy and regulatory institutions and policymakers comprised the second most frequent (42%) category of impact beneficiaries. This category of impact beneficiary included all levels of government, including in the improvement of government programmes or services.

Finally, 'industry or business' was also a relatively frequent category of beneficiary from research data-driven impact (38%), gaining improved business outcomes, profits, enhanced service and product quality and other benefits.

An example of benefits delivered to industry comes from research conducted at the University of Portsmouth, working with the UK SME Flight Data Services Ltd. (FDS), a company specialising in airline data analysis. The case study showed that "FDS Ltd had contracts with 8 airline companies covering more than 300,000 flights annually, and was required to review daily around 10 gigabytes of data." The researchers optimised the analytic process through artificial intelligence and advanced pattern recognition techniques developed using research data. This had a direct impact on FDS's business performance, saving "as much as 4,000 man-hours per year" and "freeing up of skilled staff who were then re-assigned".

[4] More than one beneficiary would be identified for each impact during the data analysis.

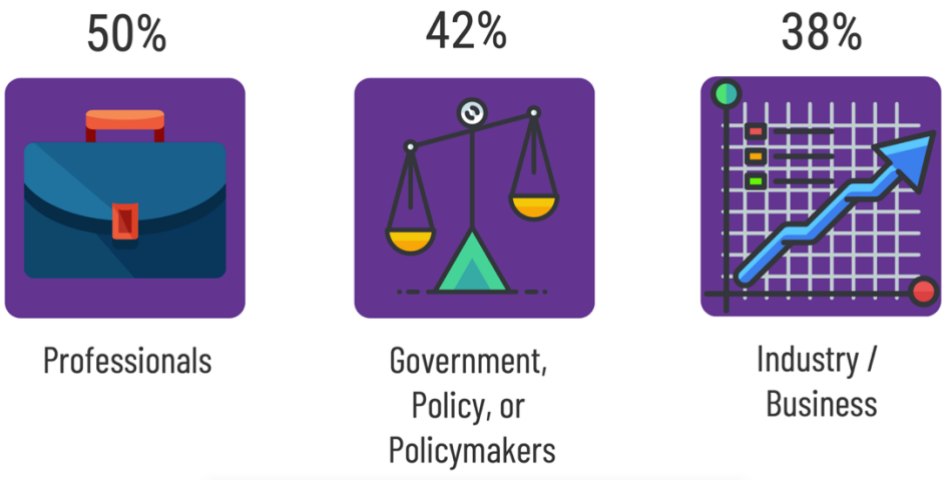


Figure 1. Three most prevalent types of beneficiaries in the analysed impact cases

4.2.1 The link between how impact is developed and who it benefits

The following table includes only variables with statistically significant negative and positive correlations. The closer the number is to 1 or -1, the stronger the correlation, with the sign indicating either a positive or negative correlation. The analysis did not reveal a strong correlation between most variables. The following points summarise statistically significant patterns identified through correlation analysis:

- 'Searchable database' tends to be used with the 'general public' (r = .22), while enhancing institutional processes / methods is not (r = -.26).
- 'Analytic Software or Methods' (r = .23) and 'improved institutional processes / methods' (r = .32) are used more to develop impact with 'industry / business'.
- 'Sharing of raw data' is more often an impact development pathway with 'environmental impacts' (r = .2).

5 UNIQUENESS OF IMPACT PATHWAYS ASSOCIATED WITH RESEARCH DATA

This section reports on findings regarding how unique impact pathways offered by research data may be. This was the final step in the content analysis that was conducted for this study: The analyst assessed, based on the judgment and the information available within the impact case study, whether the identified impact could have developed without the underlying research data. That is, was there an alternative pathway to impact outside of the research data visible in the impact case narrative content?

5.1 THE IMPACT COUNTER-FACTUAL ANALYSIS

As information about the uniqueness of research data-linked impact pathways were not inherent to the REF case study narratives, analysts had to rely on their judgement and the available information within the case to answer 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 question to the available data revealed a striking finding: Analysts judged 97% of identified impacts to be reliant on research data. That is, the identified impacts could not have been developed without the supporting research data.

	Frequency	Percentage
Impact counter-factual	439	97%

Table 7. Prevalence of research data linked impact

5.2 EXAMPLES OF UNIQUE IMPACT PATHWAYS FOR RESEARCH DATA

To illustrate how research data can provide a unique pathway to achieving an identified impact, two examples of impacts that were judged to be uniquely linked to research data are offered below.

Example 1

Researchers at the Durham University Physics Department discovered a method to uniquely identify objects by the nanoscale structure of their materials. Among other applications, the technology known as Laser Surface Authentification (LSA) allows to uniquely identify packages of “luxury goods and pharmaceutical products”. The process “has been described as being like a fingerprint or DNA sequence for the item, [because it] is unique for every document, card and carton and can be used to uniquely and unambiguously identify the item.” Without the underlying research data, these material properties would likely not have been discovered and the technology could not have been developed.

Example 2

An impact case focused on research in the area of climate reconstruction, developing “the longest established dataset having […] combined global land and marine temperature reconstructions”.

> These “instrumental temperature series are widely used as a cornerstone in nearly all discussions about climate change, often in the context of reconstructed changes in climate over the last one to two thousand years.”

Their research based on this data was fundamental to policymaking about climate change, providing the basis of the “2 degree policy”, which governments worldwide follow to limit “global warming to no more than 2 degrees Celsius above the pre-industrial average temperature”.

5.3 LIMITATIONS

The very 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. This means that 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 REF impact case studies specifically because of multi-dimensional pathways to impact making causal claims more tenuous.

6 CONCLUSIONS AND RECOMMENDATIONS

Impact is rarely delivered directly through research data. Instead it is developed through different means, as the analysis shows. ‘Improved Institutional Processes / Methods’ (40%), ‘Reports or static information’ (32%) and ‘Analytic Software or Methods’ (26%) were found to be the most frequently employed ways of developing impact from research data in the analysed REF impact case studies. 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.

Immediate beneficiaries of research data linked impact are often ‘Professionals’ (50%), ‘Government, Policy, or Policymakers’ (42%) and ‘Industry / Business’ (38%), which in turn develop insights, services, products and policies that lead to broader public impacts. Research data are therefore playing an upstream role within the research and innovation system. 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 the scope of academia requires not only traditional academic research skills, but also capabilities in public communication, entrepreneurship and boundary-crossing.

Furthermore, impact development can be facilitated through closer links between government, industry and researchers, capacity building and funding both for researchers to effectively use research data for developing impact and for potential beneficiaries to establish links with researchers and make research data available in usable formats.

The analysis revealed that research data were a basis for impact in a surprisingly high proportion of impact instances (97%). 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.

7 ACKNOWLEDGEMENTS

The Institute for Methods Innovation (IMI) would like to express deep appreciation for the time, consideration and guidance of Dr Paul Wong, Professor Mark Reed and the ARDC Advisory Group. In addition, we are grateful for the ARDC's interest in developing non-academic impacts from research data. The project also benefited from the contributions of the UK REF's impact case study authors and the open publication of these case studies. Finally, the researchers contributing to this project include: Dr Eric A. Jensen, Dr Jessica Norberto, Dr Benjamin Smith, Dr Aaron Jensen, Lars Lorenz, Christian Moll and Professor Mark Reed.

8 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.

8.1 RESEARCH DESIGN

The planned research evaluates the accounts contained in 2014 impact cases for the UK's Research Excellence Framework (REF). The planned research focuses on the content of REF2014 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.

8.2 INTERCODER RELIABILITY ANALYSIS

The 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. In accordance with this, 10% of the cases analysed by at least two analysts were randomly selected and tested for inter-coder reliability using Krippendorff's Alpha (or 'Kalpha'). From those results, 52 units were analysed. Variables showing '1.0' in the Kalpha table means that there are no disagreements by the two analysts.

Variable Name	**Alpha**
Impact Category	0.9288
Searchable Database	1.0
Report or Static Information	0.9618
Mobile App	1.0
Analytic Software or Methods	0.9478
Improved Institutional Processes / Methods	1.0
Sharing of Raw Data	1.0
Sharing of Tech / Software	1.0
Other Impact Instrument	1.0
Unclear / Uncertain	1.0
Impact Pathway	1.0
General Public	0.9362

Specific Public	0.9179
Media	1.0
Professionals	0.8836
Government, Policy, or Policymakers	1.0
Industry / Business	1.0
Other Organization	0.8774
Natural Environment	1.0
Unclear / Uncertain	1.0
Impact Counter-Factual	1.0

Table 8. Krippendorff's Alpha

The results show that there were very good inter-coder reliability scores across the variables (all above .8 Kalpha, which is the established benchmark for good reliability). All differences were resolved by the lead researcher.

8.2.1 Corrective Measures Employed during the Data Analysis

The first round of data analysis involved the use of full impact narratives from the Research Excellence Framework. This approach resulted in relatively low inter-coder reliability scores on some variables. It was determined that this was because the impact narratives were so long and contained so much extraneous detail irrelevant to the present research. For this reason, the new approach of extracting the relevant content from the cases in advance to put in front of coders was devised and implemented. This new approach yielded much higher inter-coder reliability scores.