Researchfish: A forward look

Challenges and opportunities for using Researchfish to support research assessment

Saba Hinrichs, Erin Montague, Jonathan Grant

November 2015

Researchfish: A forward look

Challenges and opportunities for using Researchfish to support research assessment

Research Report 2015/11

The Policy Institute at King's College London

November 2015

Preface

This report, commissioned by Researchfish, explores current challenges as well as future opportunities for Researchfish and its user community. The study aims to critically appraise Researchfish as a tool for assessing the outputs, outcomes and impact of research, identify the challenges faced by its users and identify future opportunities for Researchfish and its user community as an enabler for impact assessment.

For more information about this report, please contact:

Professor Jonathan Grant
Director of the Policy Institute at King's & Professor of Public Policy
King's College London
First Floor
Virginia Woolf Building
22 Kingsway
London WC2B 6LE

Tel: + 44 (0)20 7848 1742
Email: jonathan.grant@kcl.ac.uk

Contents

Figures, tables and boxes

Tables

Boxes

Executive summary

Researchfish is an online platform that enables research funders to capture and track the impact of their investments as well as enabling researchers to log the outcomes of their work. Given the large uptake of Researchfish by Research Councils and medical research charities, and its near universal coverage in the UK, it is important to examine how to maximise its impact for funders, researchers and other stakeholders interested in using the information contained in the underlying dataset.

This report, commissioned by Researchfish, explores current challenges as well as future opportunities for Researchfish and its user community. Based on interviews with 15 key individuals from research funding organisations and higher education institutions (HEIs), a brief review of the available literature and an analysis of the data collected through Researchfish, the study aims to:

- critically appraise Researchfish as a tool for assessing the outputs, outcomes and impact of research
- identify the challenges faced by its users
- identify future opportunities for Researchfish and its user community as an enabler for impact assessment.

The Researchfish online platform is designed to enable researchers to report the outcomes of their work across multiple funders, to reuse their data for their own use and have control over who sees and accesses the data. It is not intended to provide universities or funders with the tools to manage either research activity or grants, as it is essentially a data collection service for funders. As of 2015, Researchfish captures information on behalf of 79 research funders (74 registered in the UK, 5 overseas), with 63,965 principal investigators supported by the awards of these funders. Since October 2008, output, outcome and impact data has been provided by researchers, representing more than 7 years of research activity. The total value of all awards tracked in Researchfish is just under £40 billion, across all research disciplines.

Research funders and principal investigators have invested a lot of time, effort and money in collecting and inputting data into Researchfish – some of which is already being used by funders for internal analysis purposes. The key finding from our analysis is that as much effort now needs to be invested in maximising the value of the data for the wider research community.

The data collected via Researchfish is particularly important because it can contribute to four activities that characterise research impact assessment: **advocacy** for research funding, **accountability** to the funders of research, **analysis** to understand what works in research and leads to impact, and the **allocation** of future research funding.

To achieve benefits in all these areas, and maximise the value of the data collected in Researchfish, we have identified four elements that the research community needs to develop further:

(i)   data analysis and sharing
(ii)   analytical capability and capacity
(iii)   data integrity
(iv)   data connectivity within the 'research ecosystem'

Data sharing between funders is important because it can enable informed comparisons of the types of research outputs, outcomes and impact by funder, specific programme grants, and research institutions

The Researchfish dataset currently contains over one million reports of outputs (ie research output items entered by researchers), spanning over 7 years of research activity in the UK and elsewhere. These data can be analysed in a number of different ways - for single research funders interested in their outputs, outcomes and impacts; in aggregate form, illustrating the contribution of the research community as a whole; or comparatively across research funders (or research institutions) to better understand what works in research funding.

Capacity and capability of research funders and HEIs will need to be developed to analyse the data in Researchfish

One of the biggest challenges identified in our study is the capacity and capability of research funders and HEIs to analyse the data in Researchfish. While some funders are already using the Researchfish dataset to produce narratives, track individual projects and respond to questions from stakeholders, further analysis could be undertaken with the right skill set and capacity. The skills associated with research evaluation have traditionally been relatively niche, often driven by academic interests rather than practitioner needs. This is changing in the UK and internationally with the increased focus on research impact, but still represents a significant challenge. Larger funders have in-house evaluation and analysis teams to produce analytical reports, but it has taken them considerable time to develop this capacity and capability, and such resources may not be available to smaller funders. Solutions could include providing training for research funders and administrators, forming a consortium for analysing data across funders, or engaging a third party to analyse the data regularly.

Data integrity and quality will need to be continually improved and maintained

For research impact analyses to have value, the data itself needs to be of high quality. A recurrent theme from our study was the completeness and accuracy of the data inputted into Researchfish. Researchfish has introduced initiatives to validate and correct the data where necessary and this needs to be communicated to the principal investigators who input the data – to increase awareness about how the data is being used and the importance of entering outputs accurately and extensively.

Better connectivity of data within the research ecosystem contributes to data validation and maximises the potential of the dataset for analysis

To maximise the value of the data collected and ensure that it is shared across the research community, it is imperative that there is connectivity between systems, with open data sharing and the avoidance of double data entry. This requires data sharing and interoperability between systems as well as agreement about data standards. Work is underway to share information with publication datasets (such as Web of Science, PubMed and Scopus) and patent datasets, which will enable the easier validation of data when entered by principal investigators. A unique identifier of researchers is

being added through ORCID, enabling the connection of data that has researcher specific data and thereby allowing connectivity with external databases.

Future opportunities

The data contained within Researchfish has never before been available in this format, scale and level of comprehensiveness. New approaches and capabilities are needed to maximise its use. We have identified four key opportunities for the research community to take this forward:

1. Provide a 'safe harbour' to encourage data sharing across multiple funders - this could include engagement activities or communicating more widely the analytical power that is available to all funders in sharing data.
2. Develop opportunities for building capacity and capability - this could be achieved by investing in in-house capability for data analysis and forming a consortium for analysing data across funders, or engaging an external party to analyse the data regularly.
3. Continue efforts to improve data integrity, while raising awareness among principal investigators of how data are being used to encourage strong compliance, and supporting them when facing difficulties in entering data.
4. Explore connectivity with other parts of the research ecosystem and information systems, and communicating the analytical power of the dataset with pooled, high quality data.

Acknowledgements

1 | Background

1 | Background

Introduction

Researchfish is an online platform that enables research funders to capture and track the impact of their investments, and also enables researchers to log the outcomes of their work.[1] Given the large uptake of Researchfish by Research Councils and medical research charities, and its near universal coverage in the UK, it is important to examine how to maximise its impact for funders, researchers and other stakeholders interested in using the information contained in the underlying dataset.

This report, commissioned by Researchfish, explores current challenges as well as future opportunities for Researchfish and its user community. The study aimed to:

- critically appraise Researchfish as a tool for assessing the outputs, outcomes and impact of research
- identify the challenges faced by its users
- identify future opportunities for Researchfish and its user community as an enabler for impact assessment.

The report is based on:

- interviews with 15 key individuals from research funders and research organisations (in this case higher education institutions (HEIs))
- a detailed documentary review of literature and other sources in the public domain
- analysis of data collected through Researchfish
- our expertise and experience in research impact assessment.

The remainder of this chapter provides a more detailed description of Researchfish and its functionality, including an outline of how the tool was developed, the uses of the underlying dataset and a detailed description of the methodological approach adopted for this work. Chapter 2 then addresses the key themes that arose from our assessment of the current challenges and opportunities of Researchfish as a tool for collecting the outputs, outcomes and impact of research. Chapter 3 summarises the key observations and recommendations for the future.

Terminology for the outputs, outcomes and impact of research

The terms 'outputs', 'outcomes' and 'impact' of research have been used and defined differently according to different funders or research process frameworks. One prominent approach to model research processes is the payback model, which identifies the various stages following research inputs through to final outcomes.[2] These stages are the inputs to research, the research process, primary outputs from the research, dissemination leading to secondary outputs such as policymaking and product development, adoption by practitioners and the public, and final outcomes. The term 'impact' is currently used widely in research, especially with the inclusion of non-academic impact as part of the latest Research Excellence Framework (REF).[a]

---

a   REF 2014 is a process for assessing the quality of research in UK HEIs. It replaced the Research Assessment Exercise (RAE), which occurred on a (near-) quinquennial basis since 1986. The results were published on 18 December 2014. (See http://www.ref.ac.uk/ and http://results.ref.ac.uk/ for further information).

Within the Researchfish online interface, all individual entries are labeled as 'outputs' of research, including academic and non-academic outputs and any wider outcomes that may be considered 'impact'. Throughout this report we only make a distinction between the academic outputs of research (eg primarily publications) and the non-academic wider outcomes of research, which may include what could be considered 'impact' of research (Figure 1). This simplified overview is not meant to imply a linear process of research, as there are iterative processes occurring which are not intended to be captured in the Researchfish database. The value of the database is in providing a registry or indexing function to collect outputs, outcomes and impact, which can then be selected and used to find more information or for analysis in aggregate form.

Figure 1: Simplified view of outputs, outcomes and impact of research
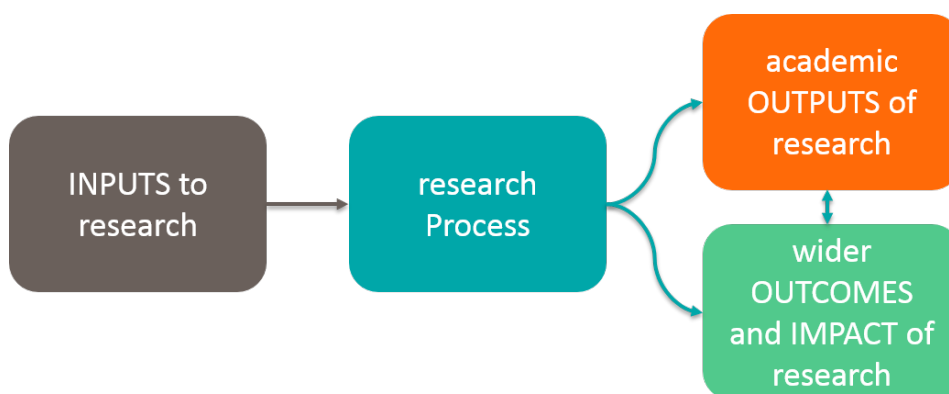
About Researchfish

Researchfish is an online platform designed to enable researchers to report the outcomes of their work across multiple funders, to re-use their data for their own purposes and to have control over who sees and accesses the data. However, Researchfish is not a type of Current Research Information System (CRIS) or grant management system as it is not intended to provide research organisations[b] or funders with the tools to manage research activity or grants. Researchfish is essentially a data collection service for funders. Figure 2 illustrates the relationship and information flows (via the Researchfish platform) between funding institutions and principal investigators. Funding is received by principal investigators who report back on research outputs to the funders via Researchfish and are directly accountable to the funders to provide this data. The data inputted into Researchfish is owned by the principal investigator, but under the terms and conditions of the research grant it is transferred to the research funder who can share the data (eg back to universities and other data platforms).

---

b   We refer to research organisations throughout this report as institutions that conduct research and receive awards from funders, which includes HEIs and any other organisation or host institution holding a grant from a funder that is signed up to Researchfish.

Figure 2: Relationship and information flows between funders and principal investigators via Researchfish
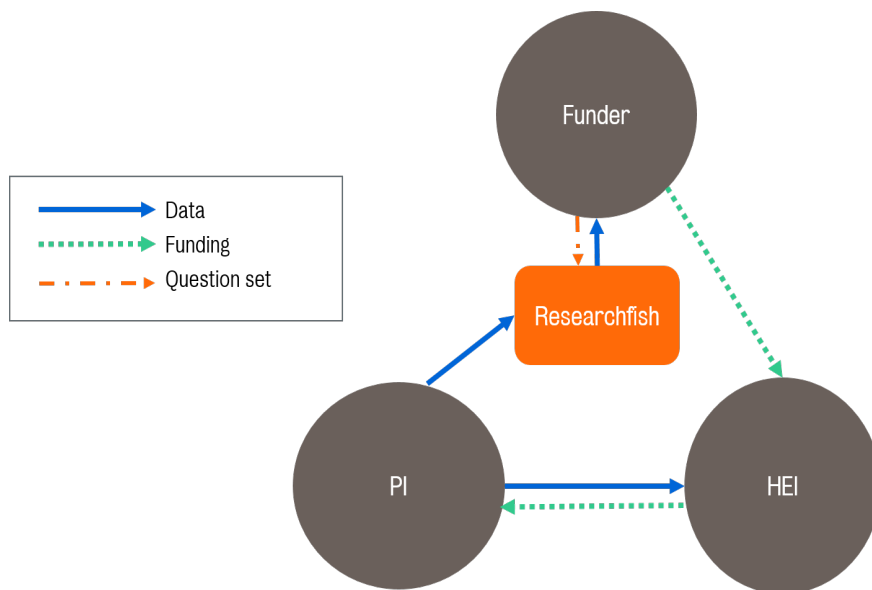
Research outputs (and outcomes and impact) are gathered through a 'question set' developed by funding institutions through a consultative process. This set of 16 questions contains 175 sub-questions as illustrated in Figure 3 (the full set of questions are available in Annex A). A researcher, or one of their delegates, can add, edit and delete entries, and crucially, attribute entries to research grants and awards. This collation and attribution of research outputs and outcomes serves a number of purposes. Research funders can capture a range of data that have been submitted by the researchers they fund – from publications, policy impact to products and interventions – enabling them to to evaluate the impact of their research funding by various units of assessment (eg disciplinary focus, research funding mechanism, host institution etc). Such evaluations strengthen accountability to the taxpayer and donor communities, and can be used to assess the effectiveness of different aspects of research funding (we elaborate on these concepts further in Chapter 2). The goal is to provide funders with agile ways to discover how work across their research portfolio is progressing and what it is producing (ie knowledge, leverage, connections etc). It also allows funders to quickly search larger portfolios and evaluate progress, productivity and quality. For example, the Medical Research Council (MRC) has published a series of reports[3] and RCUK has produced a set of reports using data captured by Researchfish.[4] The tool also allows researchers to build their academic CVs based on their research outputs from specific funders.

As of 2015, Researchfish captures information on behalf of 79 research funders (74 registered in the UK, 5 overseas) with 63,965 principal investigators supported by the awards of these funders. Since October 2008, output, outcome and impact data has been provided, representing more than 7 years of research activity.

The total value of all awards tracked in Researchfish is just under £40 billion[c] across all research disciplines (although as the biomedical and health science research funders were early adopters of the system, the majority of awards being tracked are in these fields - Figure 4).

---

c   The figure of £40 billion is based on the sum of the value of the awards recorded in the system (adding up to £24 billion) plus an estimated total of other awards that do not have exact values recorded.

Figure 3: Question set categories in Researchfish and examples of detailed sub questions for 'Publications' and 'Further funding'
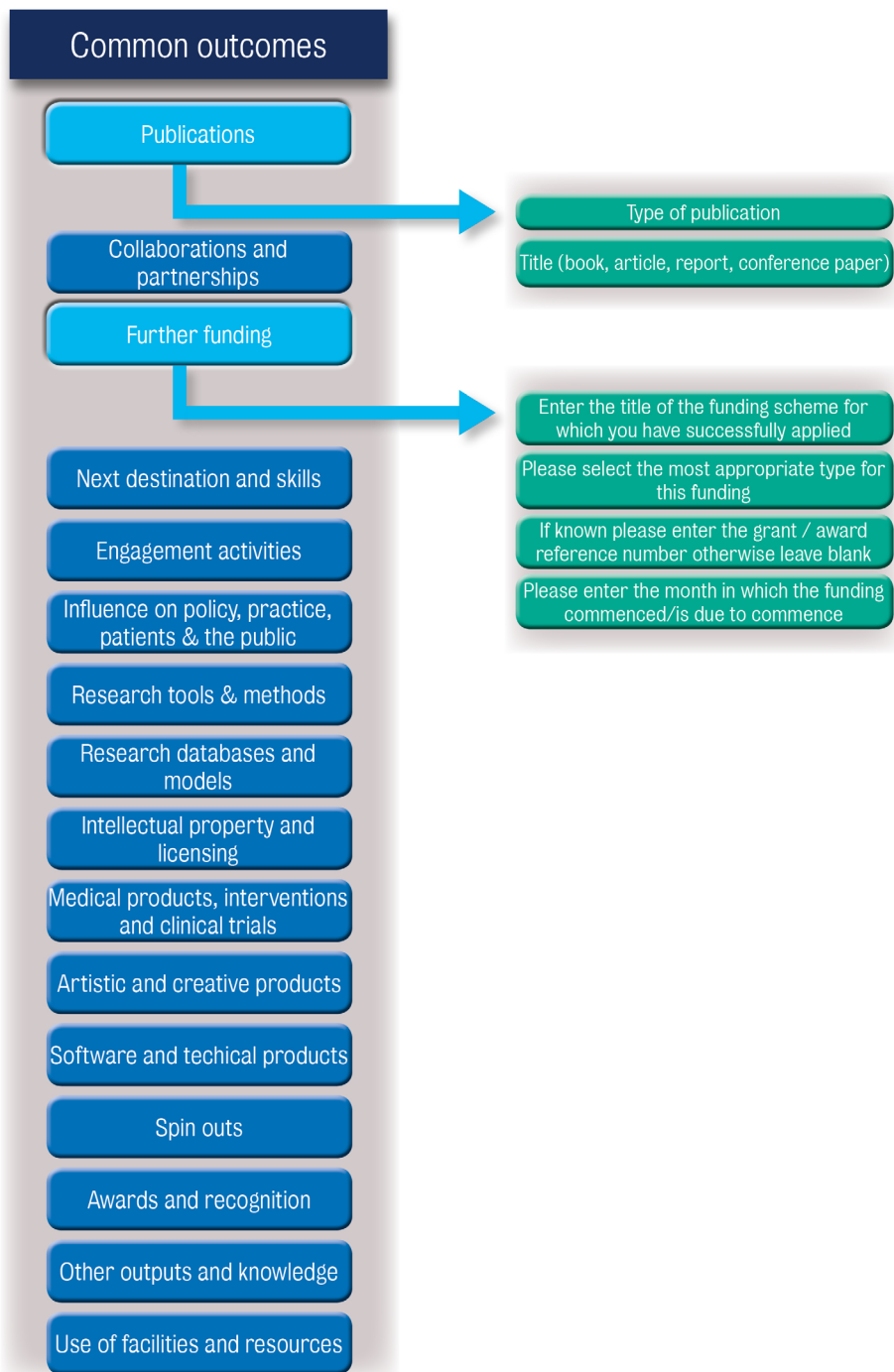
Figure 4: Estimate of the distribution of award funding per discipline (total = c £40 billion)[c]
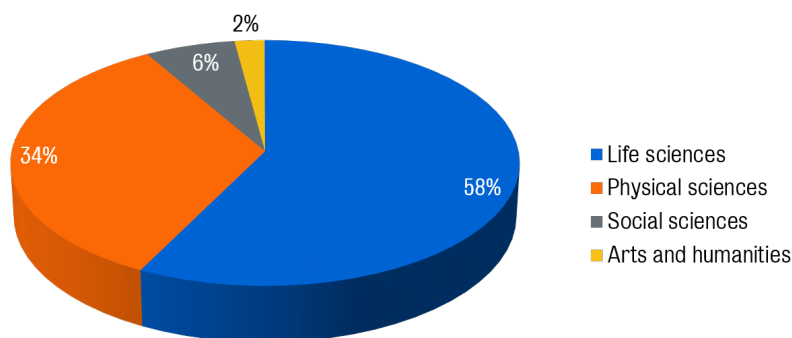
History of Researchfish

The antecedent of Researchfish was a tool developed between 2006 and 2008 by researchers at RAND Europe[d] on behalf of Arthritis Research Campaign (ARC), now Arthritis Research UK. ARC wanted to 'develop a new survey system that would provide an overview of the impacts of research ARC funded through an information gathering tool (survey instrument) that would be quick and easy for researchers to complete'.[5] As part of this work a set of ideal characteristics that the tool should aim to fulfill were identified (see Box 1). The tool first became operational in April 2008 using Selectsurvey and was named the Research Assessment Impact Scoring System (RAISS). Figure 5 provides an overview of the key milestones in the history and development of Researchfish.

During the course of the ARC project, the MRC identified a need to improve the evidence of progress across its portfolio. A joint MRC/Ernst and Young review of MRC governance recommended in 2007 that the MRC should establish a dedicated evaluation programme to provide a focus for gathering, understanding and communicating research progress, productivity and quality. Dr Ian Viney was appointed to lead the MRC evaluation programme at the end of 2007. At that time, the MRC largely relied upon researchers volunteering details of achievements arising from MRC funded work each year, a process which identified a few hundred reports, mostly research publications, and clearly could not do justice to the impact of MRC funded work.

The MRC was made aware of the RAISS tool and engaged RAND Europe as consultants to the MRC evaluation programme[6]. However, the RAISS tool only scored the presence or absence of output types and the MRC was interested in capturing addidtional qualitative details of research output as well as quantitative evidence. The MRC built on the structure of the RAISS tool and developed the questions to capture more detail to investigate the influence that particular outputs had exerted on developing impacts. The new online, systematic, structured and prospective approach would need to replace MRC final grant reports and the annual

---

d   Led by Dr Steven Wooding, RAND Europe and including Jonathan Grant, on of the authors of this paper.

Box 1: Ideal characteristics for a research output data collection tool (as developed by Wooding et al, 2009)

Capture the full range of benefits

This should include the benefits and impacts beyond publications and research qualifications.

Aggregation

The survey should allow the impacts of many grants to be aggregated, in order to provide an impression of the overall impact of a group of grants. At the same time, it has to allow for the impact of very different types to be kept apart – for example, the production of knowledge and influence on health policy. This would allow the different strengths of different types of research to be explored.

Valuation

The survey should provide a way of considering the differing value of different types of impacts, ie a method of reducing a range of impacts to a common currency.

Low burden

Any survey instrument always has a burden attached, whether this is the time it takes to complete a questionnaire, or the administration costs involved. The burden will be felt only if it is disproportionate to the benefit of conducting or completing the survey. It is important to be disciplined about the information elicited: collect only what can be used, and resist the temptation to gather extraneous information simply because there are tools to do so.

Wide applicability

The instrument has to be widely applicable across all forms of research, while allowing room for some variation.

Fairness

The instrument should capture information fairly, allowing true comparisons of groups of research grants or types of research.

Timeliness

The speed with which the instrument can provide information always will be a trade off between the requirement for speed to support decision making and allowing time for the outcomes of research to develop. Where possible, a monitoring system can provide early indicators of impact.

Figure 5: Key milestones in history and development of Researchfish

April 2008 - First tool operationalised by ARC (RAISS)

2008 - MRC uses ODGT to collect data for web-based stakeholder surveys

2009-11 - Other funding agencies adopt E-val

2011 - E-val becomes a 'federated' system for research funders

2014 - The RCUK subscribe to Researchfish

2015 - First international research funders join

2007 – MRC interested in an evaluation tool to track progress of awards funded

April 2008 – ODGT is piloted

2009 - E-val is developed

2011 - E-val spins out into Researchfish Ltd.

June 2012 - Live launch of Researchfish Ltd.

Today – 90,359 awards and £40bn research funded and projects research reported via Researchfish

achievement collection exercise. MRC wanted to move away from a narrative, snap shot view of progress at the end of a grant with a biased selection of achievements, towards researchers providing quick structured feedback throughout the lifetime of the grant and after completion. This long-term follow up was considered important to capture the way that outcomes and impacts develop.

The first version of the online survey, the 'Outputs Data Gathering Tool' (ODGT), was piloted in 2008. The survey used the Achieve Forms product supported by Firmstep Ltd., already licensed to the MRC for occasional web based stakeholder surveys. MRC soon found that it lacked the infrastructure to scale the hosting of this process and so Firmstep Ltd. was engaged to develop the survey further based on the ODGT pilot and to provide hosting and technical support. The result was MRC E-Val which successfully ran data gathering exercises between 2009 and 2011. Other funding agencies took an interest in the MRC approach, with the Chief Scientist Office in Scotland and the Science and Technologies Facilities Council (STFC) also implementing MRC E-Val, followed by the Wellcome Trust adapting the approach to collect details from their grant holders.[7]

In 2011, Mark Connelly of Firmstep Ltd saw the opportunity to make E-Val into a 'federated' system whereby research funders beyond the MRC could subscribe to a platform where principal investigators provide research output and outcome data once, and that data was then attributed, or 'federated', to different research funders. The MRC saw the value in such a system to address the risk that an increasing number of diverging, separate implementations of E-Val, would multiply the burden placed on researchers and to also to open up the possibility of national and international cross funder analysis of output. The MRC agreed to the suggestion that E-Val be spun out into a new company called Researchfish Ltd in October 2011.

Since the founding of Researchfish Ltd in 2011, and the launch of the Researchfish system in June 2012, its scope and stakeholder community has expanded. From working with just six clients, including the MRC and STFC, Researchfish grew to work with other funders of biomedical and health research in the UK, including members of the Association of Medical Research Charities (AMRC)[8]. In early 2014, the UK Research Councils (RCUK) agreed to subscribe to the system and the question set was reviewed and amended to address the full range of outputs arising from all research disciplines (a process which took around eight months). With the adoption of Researchfish across the Research Councils, following a competitive tender process[9], the MRC relinquished its intellectual property in E-Val having decided to make the detail of all data fields in the question set openly available so it could be used by anyone wanting to structure the details of research output.

As of October 2015, 90,359 awards have been entered into Researchfish (Figure 6). In 2015, Researchfish agreed to work with its first two international funders – Novo Nordisk and Alberta Innovate Health Solutions - to explore opportunities in North America, Australia, Europe and elsewhere. The increase in June 2014 was due to the implementation of the system for inclusion of RCUK, and the spike in August 2015 was due inclusion of studentships by RCUK.

Figure 6: Number of awards managed through Researchfish (total awards = 90,359 as of September 2015) over time
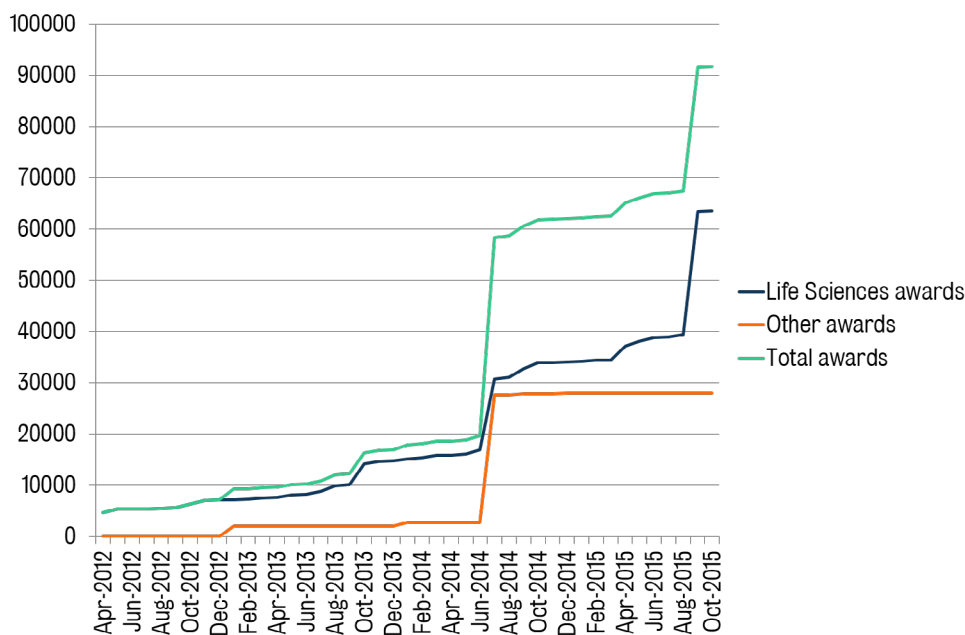
Approach

This project consisted of two activities: scoping and data analysis, and interviews to enhance our understanding and the insight accumulated.

Scoping and data analysis

Data and background information provided by Researchfish and external sources of information were analysed. The information collected included a detailed documentary review of comments about Researchfish on social media as well as literature and other sources of information available in the public domain. We also carried out an appraisal of the platform interface and analysed the dataset provided by Researchfish to provide a snapshot of the current values and outputs available to research funders. The online question set available to funders and research organisations was also reviewed to further our understanding of the user experience.

Interviews

In order to understand the challenges and opportunities of Researchfish from the user perspective, the following interviews were conducted (full list of interviewees available in Annex B):

1. Interviews with six funding institutions; one person per institution (n=6 interviews)
2. Interviews with six research organisations (HEIs). Up to two interviewees per institution were carried out - one with an academic researcher and one with a high level administrator or manager (n=9 interviewees accepted in total)

All interviews were recorded for the benefit of the research team (full interview questions can be found in Annex C).

Sampling strategy

Information on total funding and location was collected for each funder and HEI in order to stratify all organisations. Organisations were selected at random from each of the three subgroups to ensure diversity in size and location.

2 | Maximising the value of Researchfish data

2 | Maximising the value of Researchfish data

Research funders and principal investigators have invested a lot of time, effort and money in collecting and inputting data into the Researchfish data platform. The key finding from our analysis is that as much effort now needs to be invested in maximising the value of the data.

In short, there is no point in collecting the data if it is not used. Researchfish data, and all research evaluation and impact assessment data, can contribute to four aims; advocacy, accountability, analysis and allocation. Each aim has a slightly different rationale, with corresponding implications for how impact might be evidenced.

Advocacy

Research funders and providers are having to compete with other public services, and, as such, must be able to advocate the need for funding of research. Leaders within the sector must have compelling arguments to 'make the case' for research. For example, the Research Councils each publish an annual impact report which describe the ways in which they are maximising the impacts of their investments. These reports include illustrations of how their research and training has made a contribution to the economy and society.[10] The analysis of Researchfish and other similar data can support the development of these cases.

Accountability

Related to advocacy is the need for the research community to be accountable to those who fund its activities, be they tax payers or donors (as summarised by one interviewee in Box 2). Good governance dictates that the recipients of public funding should be able to provide an account of their decision making. In the context of research funding, this means that funding decisions must be made in a transparent, merit based way, and take into consideration the potential for a public benefit or social impact beyond academia.

Box 2: Interviewee 1, principal investigator

'If I was a philanthropist giving millions of pounds, I would want to see what is happening with it – and even I get frustrated [about reporting], as a researcher, when you see millions of pounds going to researchers and you wonder what is coming out of it, and you see nothing coming out of it ... And this is always the case. In a lot of things money is given, and no one follows up ...

Gone are the days when you can do what you like. Now you are publicly funded ... we want to account for what you are doing ... is it leading to some benefit to the public?...'

Analysis

The collection of research impact data supports the analysis of research policy to understand what works in research funding.[11] The 'science of science' is predicated on the ability to measure research and understand how research leads to impact, with the aim of improving the effectiveness and value for money of research funding. Knowing what works and why will inform decisions about which areas of science to invest in, determining how and who should invest and identifying the returns.[12] We know only of a few examples where Researchfish data is used in this way.[e]

Allocation

The allocation of research funding based on non-academic impact is relatively new, with the REF being the first example of its application across a research system. REF2014 assessed HEIs on the basis of the quality of research outputs, the wider impact of research and the vitality of the research environment. The impact of research was evaluated through 6,975 research impact cases studies. The use of Researchfish and other data could help HEIs and funders identify future case studies, and provide the analytical framework for developing compelling narratives.

To achieve benefits in all these areas and maximise the value of the data collected in Researchfish, we have identified four elements that the research community - including data service providers, funders, research institutions and principal investigators - need to develop (Figure 7). These four elements are:

(i)   data analysis and sharing
(ii)   analytical capability and capacity
(iii)   data integrity
(iv)   data connectivity within the 'research ecosystem'

Figure 7: Four elements for the research community to develop to maximise use of the data collected by Researchfish
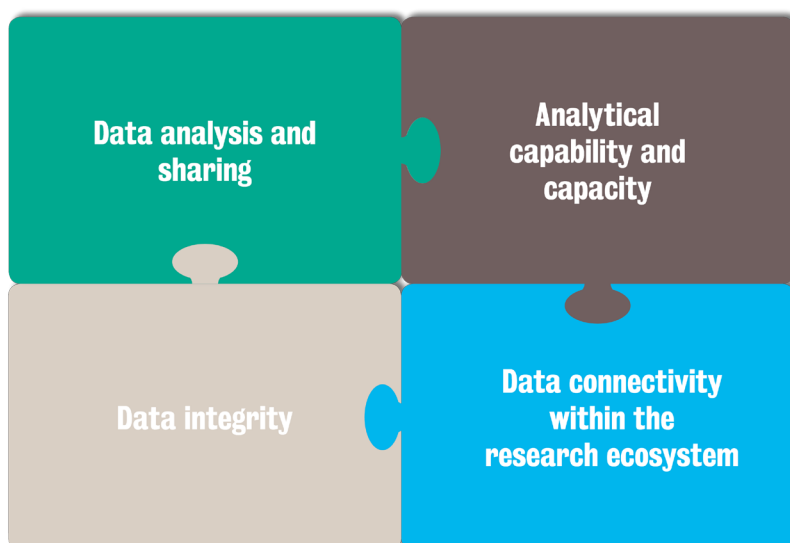
Data analysis and sharing

The Researchfish dataset currently contains over one million reports of outputs (ie research output items entered by researchers, each of which can have additional data fields) spanning more than 7 years of research activity in the UK and elsewhere. These data can be analysed in a number of different ways, for example:

- for single research funders interested in their outputs, outcomes and impacts

- in aggregate form, illustrating the contribution of the research community as a whole, or

- comparatively across research funders (or research institutions) to better understand what works in research funding.

Below we describe these types of analyses and provide examples that illustrate them.

Analysis by single research funders (and universities)

Some funders have used the Researchfish dataset for reporting purposes. For example, Cancer Research UK have produced an infographic including Researchfish data[f]. The MRC also uses Researchfish data to account for the investments it has made in the UK and overseas. In addition to a set of qualitative narratives that showcase specific research output stories, the MRC produces a dedicated quantitative report showing the patterns of research outputs in the different question set categories. With such analyses, the MRC is well placed to demonstrate the details of its funded research activity and the extent to which its awards have led to a change in policy, practice or wider adoption in society, making a strong advocacy case. Furthermore, while these examples show overall patterns, the MRC can aggregate outputs arising from different programmes, enabling it to better understand the outputs and impacts arising from specific areas of the MRC portfolio. This data can be used to evaluate specific schemes and provide evidence to support discussion on future funding strategies. Figure 8 reproduces content from the quantitative report produced by MRC in its 2013/14 annual report.[3] Figure 8a, from the section 'Policy influence', shows the distribution of types of policy influence across MRC funded investigators. Figure 8b, from the section 'Products and interventions', shows where individual research outputs are placed along the research and development life cycle.

Example A: Analysis of research outputs by different categories

Researchfish has been designed to capture information on the outputs, outcomes and impact of research that can be attributed to individual grants. Figure 9 shows one example of how the research outputs from one funder can be viewed by individual awards over time. In total, this funder currently has 227,530 individual output data pieces. From this example we can see that dissemination outputs have increased over the past five years, while awards/recognition, collaborations and further funding have leveled off.

   This type of analysis enables funders to understand and analyse the nature of the outputs from particular funded programmes and could help them to decide on future allocation. In this example, the individual awards, project name and particular research programme have not been shown. There could be an opportunity to categorise these awards, and then analyse the nature of the outputs produced per grant, discipline or specific research area of interest (eg health research categories). One of the funders we spoke to noted that their funding strategy had been influenced by examining the data in Researchfish - a funding stream which they had considered terminating was kept on their portfolio when they noted the many non-academic outputs stemming from it. Not only did they decide to keep the funding stream active, but they also promoted this funding stream further and encouraged applicants to apply.

---

f   CRUK infographic. https://prezi.com/1iyurtkesyrh/researchfish-infographic-for-researchers/ (accessed 15 October 2015).

Figure 8: Examples of quantitative analyses of MRC research outputs from its 2013/2014 report[3] (Quantitative report shown in excerpt)

8a)

- Citation in clinical guidelines
- Citation in clinical reviews
- Citation in other policy documents
- Citation in systematic reviews
- Gave evidence to a government review
- Influenced training of practitioners or researchers
- Membership of a guideline committee
- Participation in an advisory committee
- Participation in a national consultation
- Implementation circular/rapid advice/letter to e.g. Ministry of Health

8b)

Analysis of Researchfish data in aggregate form

As illustrated in Table 1, the majority (87 per cent) of researchers from the life sciences recorded in Researchfish have only one funder, while the rest have have two or more awards. Table 2 shows 43 per cent of researchers holding two or more awards. Diversity of funding is becoming an increasing pattern across research organisations[13], thereby increasing the need and potential for analysing the outputs, outcomes and impact of research activity in aggregate, as illustrated in the following examples.

Example B: Patterns of output types across awards and funders

General patterns of research activity outputs can be observed in aggregate form. The heat map in Figure 10 takes the top 20 awards with the most research output data pieces and shows the relative distribution of these outputs by type for each award. As expected, publications are predominantly the main form of output, but research materials and collaborations also form some of the main contributions of researchers. Using this heat map, a closer look can be taken at the less dominant outputs, such as 'creative products', of which only a handful of awards have produced. In Figure 10, the awards (labelled anonymously 1-20) are ordered by total number of research outputs, but the order could be changed to types of grants or disciplines to help understand the nature of outputs of each. If such data was available across funders, this comparison

Figure 9: Sample graph to show distribution of research outputs per award for one funder over time, by output type

Table 1: Number of funders awarding grants to researchers using Researchfish (UK only)

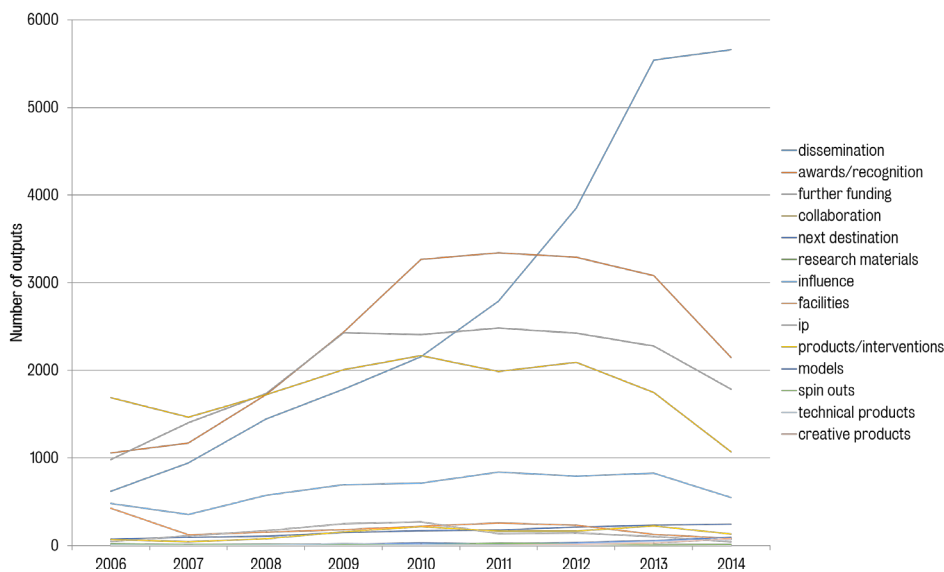|	Number of researchers	Percentage of total
No. of researchers with 1 funder	26,483	87%
No. of researchers with 2 funders	3,044	10%
No. of researchers with 3 funders	510	2%
No of researchers with 4 funders	81	<1%
No. of researchers with 5 funders	16	<1%
No. of researchers with 6 funders	2	<1%
No. of researchers with 7 funders	1	<1%
Total	30,137	

Table 2: Number of awards held by researchers using Researchfish (UK only)

	Number of researchers	Percentage of total
No. of researchers with 1 award	13,618	57%
No. of researchers with 2 awards	4,817	20%
No. of researchers with 3 awards	2,238	9%
No. of researchers with 4 awards	1,167	5%
No. of researchers with 5 awards	740	3%
No. of researchers with 6 awards	452	2%
No. of researchers with >7 awards	1,017	4%
Total	24,049	

could be made for different types of funders, for example life sciences, engineering and natural sciences, social sciences, thereby demonstrating the larger concentrations of output types in each discipline.

Example C: Analysis of non-academic impact narratives

Two of the funders interviewed viewed the Researchfish structure for output information positively, noting how it aligned with their existing frameworks for capturing impact (eg modelled through the payback framework or otherwise, as illustrated in Box 3). In particular they highlighted the potential to extract narratives that describe impact from Researchfish data.

Research administrators may be able to use Researchfish data to identify potentially interesting case studies from which to draw qualitative narratives. From our conversations with funders, there is already great value in using Researchfish to identify such narratives, and the tool has been used for this purpose when producing reports. The qualitative detail and narrative text can be as important as the quantitative data for the purposes of reporting and demonstrating outputs of research.

There is also the potential to run analyses on text based data in aggregate form by using, for example, the text from the 'other outputs' category from the question set database. More sophisticated text mining techniques could be used if more narrative text data was entered into Researchfish.

Comparative analysis across research funders (and universities)

In order to maximise the value of the data housed in Researchfish, it would be necessary to compare performance across research funders and universities. There

Figure 10: Heat map showing distribution and proportions of different output types

	publications	dissemination	awards/recognition	further funding	collaboration	next destination	research materials	influence	facilities	IP	products/interventions	models	spin-outs	technical products	creative products
Award 1	42%	24%	6%	8%	8%	6%	1%	3%	1%	0%	0%	1%	0%	0%	0%
Award 2	52%	9%	8%	16%	6%	0%	1%	7%	0%	0%	0%	0%	0%	0%	0%
Award 3	43%	6%	4%	5%	36%	3%	1%	1%	1%	0%	0%	0%	0%	0%	0%
Award 4	26%	14%	6%	12%	16%	7%	5%	3%	3%	8%	0%	0%	0%	0%	0%
Award 5	46%	48%	4%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Award 6	76%	6%	1%	11%	3%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%
Award 7	11%	5%	1%	0%	5%	3%	73%	2%	0%	0%	0%	0%	0%	0%	0%
Award 8	57%	4%	6%	18%	5%	7%	1%	1%	0%	1%	0%	0%	0%	0%	0%
Award 9	66%	13%	6%	7%	5%	0%	0%	2%	1%	0%	0%	0%	0%	0%	0%
Award 10	64%	7%	6%	7%	8%	1%	1%	3%	0%	1%	0%	1%	1%	0%	1%
Award 11	64%	8%	6%	7%	8%	1%	1%	2%	0%	1%	0%	1%	1%	0%	1%
Award 12	53%	16%	13%	12%	4%	0%	1%	1%	0%	0%	0%	0%	0%	0%	0%
Award 13	78%	7%	6%	4%	1%	2%	0%	2%	0%	0%	0%	0%	0%	0%	0%
Award 14	66%	1%	2%	29%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Award 15	53%	3%	10%	26%	5%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Award 16	77%	6%	2%	1%	9%	2%	0%	3%	0%	0%	0%	0%	0%	0%	0%
Award 17	69%	9%	13%	2%	7%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Award 18	60%	5%	6%	12%	15%	0%	1%	0%	0%	1%	0%	0%	0%	0%	0%
Award 19	73%	9%	15%	1%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Box 3: Interviewee 2, research funder

'The reporting framework basically gives five categories which include: advancing knowledge (publications), informing decision making, health systems, health and social economic impact, and capacity building. We created the portal and implemented this. We had been doing this all along, so wanted a tool that aligned ... but had a better functionality and would be more interoperable. ...

When it comes to reporting, we make sure the results are presented across the five impact categories ... Another thing I want to do is the quantitative [analysis] but also the qualitative, to use Researchfish to tell the impact case narrative and pulling that out to supplement the qualitative research.'

are a number of cultural barriers preventing this from happening, but an example of this type of analysis is a comparison of the academic and non-academic outputs by research funder.

### Example D: Comparison of academic vs non-academic research outputs across funders

The question set in Researchfish allows researchers to report on a variety of outputs ranging from academic publications to wider outcomes and impact of research that can be attributed to individual grants (see Annex A for the full question set). We extracted the information held in the Researchfish database and plotted all academic outputs (in this case, publications) against the count of all non-academic outputs ie engagement activities, policy, intellectual property and spin-outs[g] (Figure 11).

Each point in Figure 11 represents a funder, and their location on the graph is based on the number of academic versus non-academic outputs entered into the system by principal investigators. Currently there are funders which have many more total outputs recorded (eg funders labelled A and B) compared to the smaller funders, so we have included an image to zoom in and show the smaller funders.

Across the whole dataset, approximately 60 per cent of the outputs reported in Researchfish are academic and 40 per cent are non-academic. Some of the funders are outliers and have disproportionly more academic or non-academic outputs. For example, one of the outliers (labelled C) has proportionately more academic than non-academic outputs, while another has the reverse (labelled D) raising questions as to what the differences are between C and D - is one more focused on basic research and the other applied? Does one have an active 'impact strategy' in place to facilitate translation? Or has the funder directed these researchers to prioritise filling out publications or other types of outputs? The answers to these and other similar questions are important to understand what works in research funding but can only be answered if C and D are willing to be identified and exchange information on their practices. This particular example in Figure 11 may simply be a reflection of the type of data entered by researchers, rather than a true picture of where the emphasis lies for each funder and their fundees.

---

g   This includes the full set of outputs entered under the following categories: Collaborations, Next Destination and Skills, Engagement Activities, Influence on Policy, Practice, Patients and the Public, Research Tools and Methods, Research Databases and Models, Intellectual Property and Licensing, Medical Products Interventions and Clinical Trials, Artistic and Creative Products, Software and Technical Products, Spin-outs, Awards and Recognition, Other Outputs and Knowledge, Use of Facilities and Resources. We only included counts for outputs that were reported from 2006, as this was the common first date for all reported outputs.

Figure 11: Sample graph to show proportion of non-academic outputs to academic outputs of funders (anonymised)
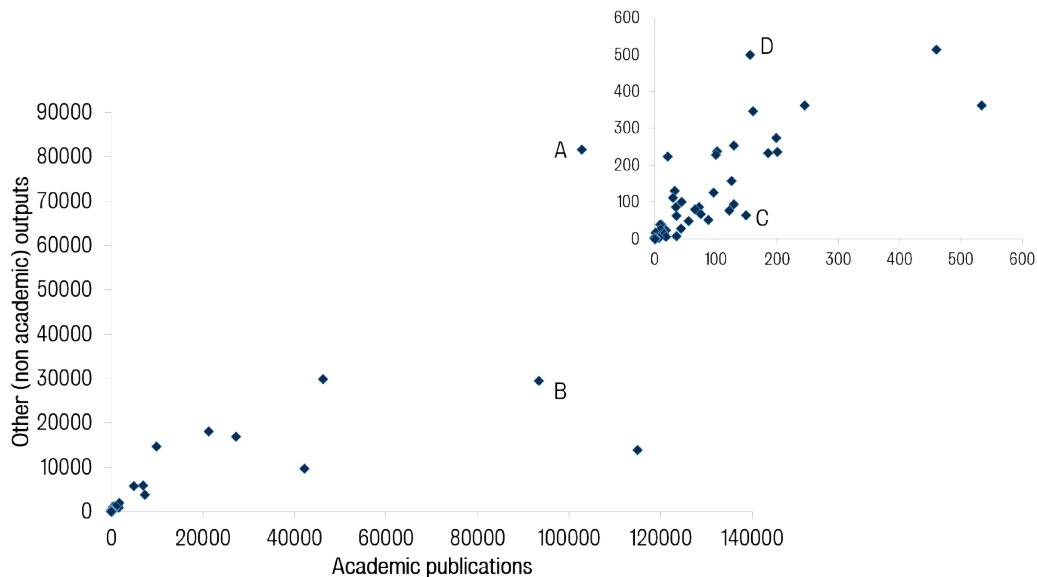
Although this is a crude analysis of the proportion of academic versus non-academic outputs, it demonstrates the potential for showing the emphasis of the types of outputs for each funder. Similar analyses could also be carried out at the research institution level, where research administrators could have an overview of the nature of outputs and compare the emphasis of academic versus non-academic outputs across their departments.

The case for data sharing

There is a clear opportunity to be gained from comparative analysis by collecting and sharing research output data – be that across funders or universities. However, it was clear from our interviews that such data sharing can still be challenging. This could be due to concerns around the other three pieces of the 'puzzle' ie data integrity, analytical capabilities and connectivity (Figure 7).

Building up trust and confidence in data systems to enable data sharing is not a new challenge. Another example is the implementation of clinical audit in the UK. The Intensive Care National Audit & Research Centre (ICNARC) was set up in 1994 in direct response to a need for high quality information (data) in the medical sector.[h] Healthcare providers at the time had an unclear picture of the effectiveness and overall treatment of patients in critical care within the UK. In 1991, the Department of Health proposed creating a national centre which would provide 'comparative audit and evaluative research' for intensive care, and as a result ICNARC was established as an independent, registered charity to monitor and evaluate intensive care at a national level. ICNARC submits a Data Analysis Report to each unit, identifying trends over time and anonymised comparative analytics to other units in the UK. To date, the database has 100 per cent participation of all adult, general critical care units in England, Wales and Northern Ireland[14], and holds 1.5 million patient cases. Growing participation in providing data for the database has taken significant effort by 249 intensive care units. The learning from this example could be used by the research community to increase widespread understanding of how to maximise value from data

---

h   Intensive Care National Audit and Research Centre (ICNARC). Website available at: https://www.icnarc.org
     (Accessed 1 5 October 2015).

sharing.

Analytical capablity and capacity

One of the biggest challenges facing the research community, identified in our interviews, is the capacity and capability of research funders (and universities) to analyse the data in Researchfish. This is a key element in being able to maximise the value of the data and deliver on the 'four As' of advocacy, accountability, analysis and allocation. This type of data has never before been available in this format, scale and level of comprehensiveness, and new approaches and capabilities will need to be supported to maximise its use.

A support system for both large and small funders

In terms of capacity, larger research funders have established internal evaluation teams to oversee the collection and reporting of Researchfish and other evaluation and research impact data. The MRC has had an evaluation team since 2007, and the entire evaluation, performance monitoring, reporting and information management and analysis team comprises of 12 staff. This has, however, taken time to develop and involved experimenting with how to make best use of the dataset. The challenge for smaller funders, who typically will have only one or two employees, is to create the capacity and capability in analysis over time. For example, 86 per cent of the value of research funding covered is accounted for by seven research funders, while seven of the smaller funders hold less than £10m worth of awards between them. For the latter it is simply not affordable to have in-house evaluation teams. One smaller funder interviewed commented on how they export all the data from Researchfish, and then, through minimal resources, clean and analyse it again for their own reporting purposes. Larger government funding bodies are also mandated to provide evidence of returns on investment, so the motivations for analysis and reporting are evident. Smaller, private funded charities may not have the same incentives for investing into more rigorous and standardised reporting. There is a need, as noted by the larger funders, to raise capacity in reporting for smaller charities, but also to provide the necessary resources.

This creates a number of challenges and opportunities for the research community. One scenario could be that Researchfish could improve the reporting aspects and functionality of the data platform, allowing subscribers to generate standard reports; while smaller research funders could pool resources and form a consortium to analyse the data. Alternatively, they could develop evaluation strategies that commit to analysing the data on a regular (eg every three to five year) basis. A third party could also enter the research community and offer an analytical service to such funders. All of the above are likely evolutions over the coming years, but a key point for research funders is the need to use the data in Researchfish if they are going to collect it. The last UK Health Research Analysis noted that although the amount invested may be minimal for smaller charities, the contribution to a specific disease area or therapeutic approach can be significant.[15]

Governing and updating the question set

All the funders subscribed to Researchfish have the opportunity to influence the question set and make suggestions for its general improvement. We interviewed both large and small funders and generally it was felt that the process of improving the question set was working well. However, we noted that the process of adding questions (especially those that may be very different from the initial medical oriented funders) required a strong voice or 'champion' during meetings, and smaller charities may not have adequate resources to send representatives to such meetings. As more funders sign up to Researchfish it may also become difficult to manage the requirements of each funder, as it can currently take months of correspondence and

meetings to update one question.

As the funders network using Researchfish increases, and more researchers are required to report, the prioritisation of questions in the question set may become necessary to allow for a useful and 'responsible' balance between analysis and collection of outputs. A further challenge may be that researchers themselves wish to contribute to the question set, or have a mechanism for demonstrating which types of outputs, outcomes and impact are important to report. In the case of REF, the narrative way of reporting impact case studies, while challenging for analytics, provided an opportunity for researchers to highlight aspects of their research that they considered important.

Making the most of Researchfish data for universities

In addition to research funders, universities could make more use of the data integrating it where appropriate into existing CRIS systems (if such capabilities were made possible in future) and using it as a resource to inform impact strategies associated with REF. One research administrator we spoke to explained how they used the information in Researchfish to find 'interesting' case studies, for the purposes of presenting success cases to the funder during a visit and the other in supporting their collection of impact case studies for the REF2014 exercise.

Many universities have recently invested significant resource in their research information management systems. Given this investment, the priority for these organisations is to encourage researchers to use their local system, in particular to address the requirements of the next REF exercise to submit outputs and to track compliance with open access mandates. Some organisations are therefore keen to explore the transfer of publication data from their CRIS to Researchfish and a pilot is under discussion to achieve this.[16]

Developing research impact assessment skills

The skills associated with research evaluation have traditionally been relatively niche, often driven by academic interests rather than practitioner needs. This is changing in the UK and internationally with the increased focus on research impact. As a result, a number of providers have developed training and development courses including, for example, the Leiden based Centre for Science and Technology Studies (CWTS) on bibliometric analysis[i], and the International School on Research Impact Assessment (ISRIA)[j]. ISRIA has held week long Schools in Spain, Canada and Qatar, as well as a regional workshop in Chile, the Netherlands and Canada (see Box 4 for more information). As demand for robust analysis of research impact data increases it will be important to continue to develop the capabilities across the research community in undertaking such analyses.

Data integrity

In our interviews a recurrent theme was the quality of the data inputted into Researchfish. For data to have value it needs to meet the highest standards of data quality. As noted in Chapter 1 (Figure 2), the data that is inputted into Researchfish is 'owned' by the principal investigator but then, under the terms and conditions of research grants, is transferred to the research funder. This research funder can then share that data, including back to universities and other data platforms as it desires.

The importance of data integrity

i   Centre for Science and Technology Studies (CWTS). Website available at: http://www.cwts.nl (Accessed 15 October 2015).

j   The ISRIA was cofounded by Jonathan Grant, one of the authors of this report. International School on Research Impact Assessment (ISRIA). Website available at: http://www.theinternationalschoolonria.com (Accessed 15 October 2015).

Box 4: The International School on Research Impact Assessment (ISRIA)

Vision

ISRIA will be a leading global collaboration for excellence and innovation in research impact assessment in all fields of science.

Mission

ISRIA will advance knowledge and build Research Impact Assessment (RIA) capacity across all fields of science through:

- Promoting understanding and optimisation of research performance

- Developing capabilities and resources on RIA

- Promoting a global community of practice and mutual learning

- Providing solutions to actors in need

- Strategically guiding nation's/institution's R&D

Goals

ISRIA aims to develop sustainable capabilities, successfully address RIA methodological innovation and ensure global recognition. Goals to address these impacts include:

Building sustainable capabilities and capacities

- Develop sustainable human capital in RIA

- Develop a community of practice for mutual learning

Advancing knowledge and methodological innovation

- Develop and apply existent and new RIA methodologies

- Improve the understanding and analytics related to RIA

- Emphasise the wider impact (non-academic and societal impact) of RIA

- Attract the attention of governments and funders

Extending global reach

- Develop international outreach and visibility

- Promote partnerships between actors

- Ensure presence in high profile events and publications.

Principles

The following are the principles that will guide ISRIA activities:

- Neutral approach to frameworks, tools etc

- Transparent, open and accessible

- Build a community of practice

- Deliver social value

- Useful, practical, feasible and cost effective

- Advance understanding of theory and practice

- Advance the evidence and practice base in RIA

The quality of the information which is collected, analysed and stored is important. There are five main attributes of data integrity; accuracy, attributable, available, complete and consistent.[17] Data quality is challenged in the current big data era specifically due to: the diversity of data sources (ie types and structures), which makes it difficult to integrate systems; the overwhelming volume of available data, which can make it difficult to judge its quality in a reasonable amount of time; the changes to data being faster than its validity and accuracy, which means processing technology must act quickly; and that currently there is not a unified and certified standard for data quality.[18] Previous literature emphasises the implications of bad quality data on the consumer[19], and highlights the importance and power of data and the potential for misuse and error in any data systems on which our health services, security and finances rely. The question of integrity of data for analysis purposes, and the ability to maintain its quality, can come down to who is accountable and who owns the data. Specifically, it has been noted that researchers and funding institutions jointly hold the responsibility for the integrity and quality of the data[20], although in the current research ecosystem where data is transferred from one database to another and pulled from different sources, there are many more instances in which data could be corrupted.

'Bad' quality data explained

Concerns around the integrity of the data contained in Researchfish arise from a number of sources, as summarised in Table 3. First within Researchfish there is legacy data that arose from E-Val. Some of these data do not directly match to the current Researchfish question set, so they will be incomplete and can appear to be of poor quality. To alert users to this issue Researchfish have just introduced a 'flag' or '▲' mark in the system. Similarly, there is a proportion of data imported from the National Institute of Health Research RAISS system in mid-2014, which did not map directly onto the Researchfish question set, leaving the appearance of poor data. Again Researchfish has just introduced a flag to highlight such data to users.

There are several steps that are taken to add to the data and check its integrity, first by Researchfish and then the funders. As part of the service provided to funders, Researchfish check geographical locations, names of companies etc and add information about the sector and country to the dataset. Researchfish also ensures that all publication outputs have unique Document Object Identifiers (DOIs) by importing such information from other datasets including PubMed Central. Research funders may check to ensure that grant numbers and such like are accurate and will query missing or erroneous data (eg large instances of further funding). Once this process is complete, and as illustrated in Table 3, Researchfish estimates that about 1% of the data within the system is 'bad', that is, incomplete or wrong.

Clearly, what is not known is how accurate the 'clean' data is, and an exercise that research funders and Researchfish may wish to undertake is an audit of a representative sample of data entered by principal investigators. One element of this was done by RCUK in looking at the comparative accuracy and integrity of publications data held in CRIS and Researchfish.[21]

Achieving compliance to increase data quality and completeness

Compliance rates of different research organisations vary. Despite difficulties in getting all academics to submit data to Researchfish, many of the funders interviewed indicated that they had a compliance rate of between 75 and 100 per cent in the last submission period. Achieving a good compliance rate can be more difficult when funders have a large international funding portfolio, where the compliance rate had previously been as little as 46 per cent, as indicated by one funder. One way of increasing compliance suggested by research administrators was to increase the facility for delegation within the system, although funders note that researchers have the more detailed understanding of the research process and associated outputs, and are therefore best placed to monitor data submitted.

Table 3: Types of poor quality data and sources

Type of poor quality data	Explanation and source of 'poor quality' data	Who does this affect?
Missing value (legacy data)	Original E-val system had a one-on-one reporting system without a portfolio, as each award was dealt with on an individual basis	Original E-val users (pre-dating Researchfish)
No value (non-mandatory data)	These errors will be visible when a principal investigator completes an outcome question that is flagged as non-mandatory when it should be mandatory	All principal investigators who enter data as non-mandatory data which should be mandatory
Pending location clean	The location data is erroneous, ie it does not match the underlying database of locations. This can be down to an incorrect spelling, ambiguous description, multiple locations in one entry or a brand new entry	Any principal investigator who enters erroneous data on location
Missing value (imported data)	Denoted by a warning triangle beside the outcome. This will be displayed when an outcome has been previously imported via an external system on behalf of an organisation by Researchfish. Any outcome displaying this triangle will not be able to be attributed against another award, and will need to be recreated before it can be reused	All outputs referring to awards imported by funders from another organisation
Missing value (bad data)	The data is bad and cannot be retrieved despite it being entered correctly within the outcome form	All data entered incorrectly

Currently, submission of data to Researchfish is the responsibility of the principal investigator holding the research grant. This ensures that researchers are personally responsible for the information held on the system. Many of the research administrators interviewed agreed that this was important for accountability reasons, but acknowledged that it can be challenging to monitor the progress of each submission.

Data connectivity within the 'research ecosystem'

The recently published Higher Education Funding Council for England (HEFCE) report on research metrics highlighted the need to improve data infrastructure that supports research information management[12]. In doing so, it acknowledged a similar tension that arose in our interviews, namely that:

'The different systems operated by HEIs, funders and publishers need to interoperate and to import and exchange data more efficiently and effectively; also definitions of research-related concepts need to be harmonised across the systems. An obvious example is the overlap between HEI institutional uses of CRISs for research management and RCUK's requirement that researchers use Researchfish for reporting, which is creating a need for the same information to be entered twice, into different systems'

Connectivity of the research ecosystem and developing a universal information management system through information and communication technology infrastructure relies on linking researchers to their research initiatives. ORCID[k] and ISNI, two person and object identifiers, complement the Researchfish information management system by better informing and sharing expertise within the research community. To date, ORCID has more than 1.4 million registered researchers and is being internationally mandated by governments into higher education policy. There are conflicting views on whether ORCID should be enforced in this way. Clear benefits of ORCID and ISNI have been identified in developing collaboration and interoperability of information management systems, while providing measurable efficiency improvements for participating universities, especially in internal data quality.[22]

The research ecosystem and its connectivity

Figure 12 illustrates the research data ecosystem from the perspective of Researchfish. Different stakeholders will have different perspectives on what this system looks like, but to maximise the value of the data collected within the research community, it is imperative that there is connectivity between systems - with open data sharing and avoidance of double data entry. It is not in the interests of anyone in the research community for a data monopoly to develop, and this requires data sharing and interoperability between systems as well as agreement about data standards (articulated by one research funders in Box 5).

At the centre of Figure 12 are research funders, principal investigators and universities. Historically, the principal investigator provided information to their employer (the HEI), which shared it with the research funder, often through ad hoc requests. The introduction of Researchfish has effectively changed the data flow: the principal investigator now shares the information directly with the research funder, and the funder is in the position to then share the information with the HEI, although in practice this may not occur regularly. Within this dynamic is a change of emphasis on the 'data unit': HEIs collate data around an employee (who may hold multiple research grants), while the new system is around a research grant (or more specifically a unique grant reference number). Many tensions raised in our interviews – and echoed in other reports[12, 21] – are associated with these changes.

---

k   ORCID is an open, non-profit, community-driven effort to create and maintain a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers. ORCID is unique in its ability to reach across disciplines, research sectors and national boundaries. It is a hub that connects researchers and research through the embedding of ORCID identifiers in key workflows, such as research profile maintenance, manuscript submissions, grant applications, and patent applications. Available at: http://orcid.org (Accessed 15 October 2015).

Figure 12: Research ecosystem from the perspective of Researchfish
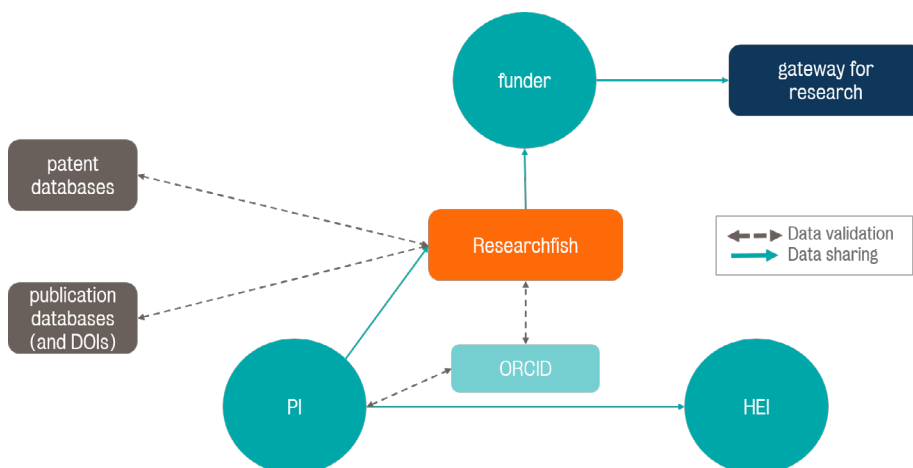
Box 5: Interviewee 3, research funder

'"Utopia" is that we have a culture of open innovation working together on collecting, measuring, analysing and optimising impact for public value, patient value – and it is a constant, adaptive system with many feedback loops that is giving value back to the beneficiaries and moving the thinking forward for our community: policymakers, decision makers etc. An ever evolving, adapting system is my "utopia". It can't simply be a tool, it has to be the use of the information for this evolving, adapting community.

"Utopia" is doing lessons learned with our community. So I am picking on the MRC – I am sure they have great lessons learned – so how do we get to the new generation, how do we start bringing this forward thinking in these multi channels? And having IT systems are part of it, but it is not all of it.'

There are existing practices of data sharing (Box 6) – for example, Researchfish pulls in information from PubMed Central (to both clean output data and include DOIs) but then also exports information to allow PubMed to include grant reference numbers in its system. Similarly, RCUK exports the Researchfish information into Gateway for Research, making it publically accessible and open. The inclusion of the unique ORCID identifier may well overcome the challenge of research funders managing at the grant level and HEIs managing at the researcher level. There is also a pilot underway to import publications data from universities via a bulk upload. Although currently limited, it will be assessed post-submission in March 2016.

Box 6: Existing connectivity flows between Researchfish and external databases

1. Europe PubMed Central: information shared both ways, which is then propagated to PubMed Central to validate publications and associated grants

2. Publications databases: information shared both ways to validate publications and associated grants (currently includes Web of Science (WOS), NASA Astrophysics Data System (NASA ADS), SCOPUS, INSPIRE, ISBN, ETHOS)

3. ORCID: information shared both ways to validate researcher identification

4. Patent databases (EPO and WPO): information is automatically refreshed when a patent moves from pending to granted

5. Currency conversion: external information flows to check currencies for funding award amounts (back office function only)

The research institutions' research outputs ecosystem

Many of the principal investigators interviewed expressed challenges around data integrity, but also highlighted the opportunity for there to be a system such as Researchfish to communicate with existing CRIS (ie grant and publication management systems) enabling both a 'push' and 'pull' of data.

The importance of this opportunity becomes clear when examining the current systems in place at a university. Figure 13 shows the various platforms that are currently in use for managing outputs and collecting information, from the perspective of the university (grant and publication management systems), the researcher and the overlap of what is also managed on the Researchfish platform. There is only a small overlap in what is captured through Researchfish, given the variety of grants and funding sources (private, industry or funders not on Researchfish) and the various routes of dissemination that researchers use for their outputs, outcomes and impacts (project-dedicated websites, other researcher online social media outlets etc).

Figure 13: Sample of exsisting channels for reporting on research outputs from the perspective of researcher
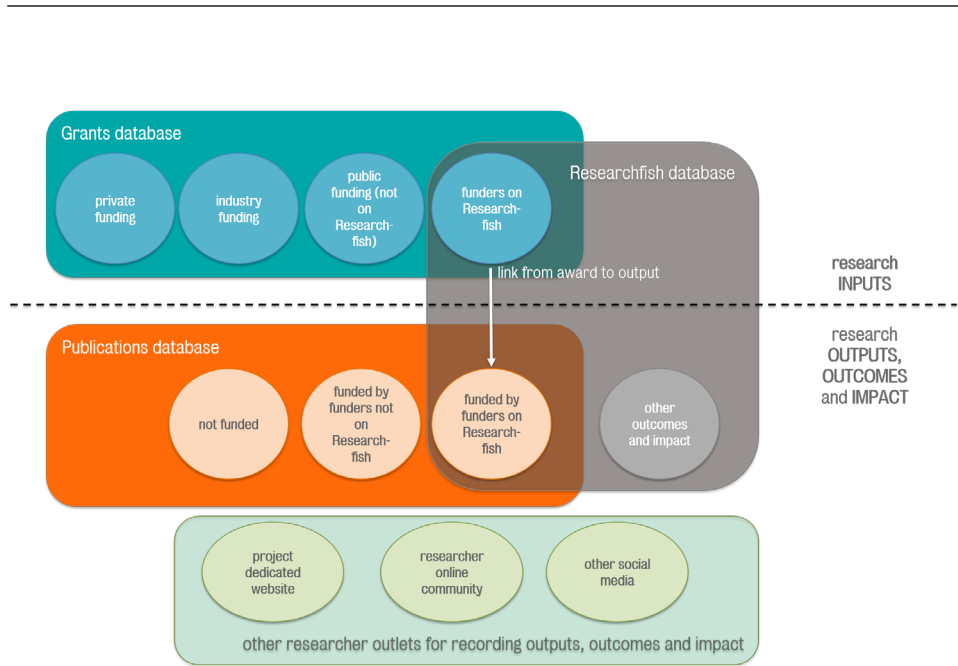
From the universities' administration perspective, there may be a requirement to hold grant information separately, as there may be institutional grants not assigned to a particular researcher, in addition to the other various funding sources not under the Researchfish network.

Researchfish was not designed to be a publication or grant management system; but as end users of the interface, researchers and their administrators perceive the platform to be something separate to their own records, and an extra task to complete to satisfy funders. The different purposes suggest the need not for one system that fits these varying purposes, but one which can 'push' and 'pull' data across various platforms. There may be an opportunity for achieving such interoperability through the use of ORCID as the connector of information. An interoperable and open data structure could also enable the development of appropriate metrics and indicators for research impact assessment and management.

3 | Next steps

3 | Next steps

Researchfish is a 'federated' system whereby research funders subscribe to the platform, researchers themselves provide the research output and outcomes data, and this information is attributed to different research funders. The advantages in the potential of the platform can be summarised in the following areas:

- Funders and research administrators have a better mechanism for collecting information on research outputs and impact, and are not reliant on text based annual reports

- Patterns of research activity can be used to benchmark across funders, research organisations nationally (and potentially globally)

- The platform has the potential to enable research impact assessment through the collection of various non-academic outputs, outcomes and impact

- The data collected improves the accountability of the impact contributed by all of the stakeholders, as well as across the research sector.

As much effort now needs to be invested in maximising the value of the data entered by principal investigators.

Key recommendations

1. Provide a 'safe harbour' to encourage data sharing across multiple funders

There are clear opportunities for comparative analyses by collecting and sharing the research output data collected across funders and universities (as examples in Chapter 2 demonstrate). From our interviews with research funders, data sharing was not seen as problematic, but as more funders join and the network expands, potential tensions may arise as funders may have preferences or governance requirements as to how such information is shared and when. This will require ensuring that funders have a trusted 'safe harbour' for data to conduct the comparative analyses, and an external partner may be required to provide such analytics. It may also mean initiatives need to be set up to encourage data sharing across multiple funders – this could include engagement activities or communicating more widely the analytical power that is available to all funders in sharing data.

2. Develop opportunities for building capacity and capability

While larger funders may continue to have in-house capability for data analysis, there are opportunities to increase capacity for smaller and new funders that join the Researchfish network. This may include providing training opportunities for research administrators through existing research impact assessment training; forming a consortium for analysing data across funders; or engaging a third party to analyse the data regularly.

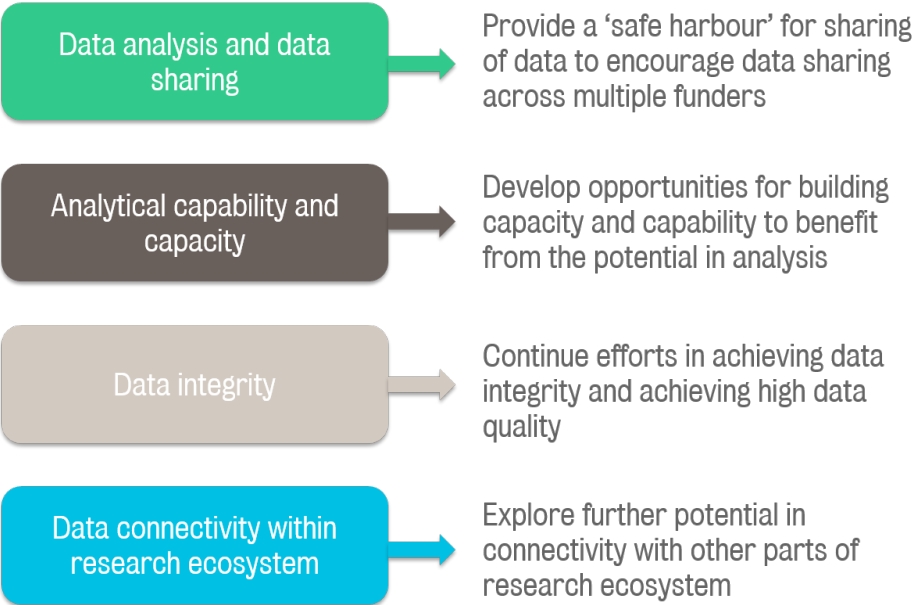### 3. Continue efforts to improve data integrity

For research impact analyses to have value, the data itself needs to be of high quality. While there are now initiatives at Researchfish to provide routes for data to be checked, validated and if necessary amended, efforts need to continue to engage and communicate with the principal investigators who input the data – to

increase awareness about how the data is being used, the importance of entering outputs accurately and extensively, as well as to continuing to support principal investigators when facing difficulties in entering data.

4. Explore connectivity with other parts of the research ecosystem

To maximise the value of the data collected within the research community, connectivity between systems is essential, with open data sharing and the avoidance of double data entry. The advantages in making Researchfish interoperable with other systems include reducing the burden on researchers by having to input data only once, enabling the 'push' and 'pull' of data to other dissemination routes or online research platforms, enabling the use of data for pre-population of future grant applications and the potential to use the data for other bespoke reports. Researchfish is already validating information with other datasets, and we encourage such efforts to continue, and to communicate the analytical power of the dataset with pooled, high quality data.

Figure 14: Key recommendations for future relating to observations

Data analysis and data sharing	Provide a 'safe harbour' for sharing of data to encourage data sharing across multiple funders
Analytical capability and capacity	Develop opportunities for building capacity and capability to benefit from the potential in analysis
Data integrity	Continue efforts in achieving data integrity and achieving high data quality
Data connectivity within research ecosystem	Explore further potential in connectivity with other parts of research ecosystem

References & annexes

References

1. Researchfish, 'What we do', 2015, http:// www.researchfish.com/about_wwd (accessed 15 October 2015).

2. Buxton, M. and Hanney, S., 'How can payback from health services research be assessed?' *Journal of Health Services and Research Policy*, January 1996, 1: 35–43.

3. Medical Research Council, 'Outputs, outcomes and impact of MRC research: 2013/14 report', 2014, http://www.mrc.ac.uk/news/publications/outputs-outcomes-and-impact-of-mrc-research-2013-14/ (accessed 15 October 2015).

4. Research Councils UK, 'Use Researchfish Data Publications 2009 – 2015', 2015, http:// http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/ UseResearchfishdatapublications2009-2015.pdf (accessed 15 October 2015).

5. Wooding, S., Nason, E., Starkey, T., Hanney, S. and Grant, J., 'Mapping the Impact: Exploring the payback of arthritis research' *Santa Monica, CA: RAND Corporation*, 2009, http://www.rand.org/content/dam/rand/pubs/ monographs/2009/RAND_MG862.pdf (accessed 15 October 2015).

6. Ismail S., Tiessen J., Wooding S., 'Strengthening Research Portfolio Evaluation at the Medical Research Council - Developing a survey for the collection of information about research outputs', *Santa Monica, CA: RAND Corporation*, 2010. http://www.rand.org/pubs/technical_reports/TR743 (accessed 15 October 2015).

7. Wellcome Trust, 'Managing a grant', 2015, http://www.wellcome.ac.uk/ Managing-a-grant/WTP051544.htm (accessed 15 October 2015).

8. Association of Medical Research Charities, 'Using Researchfish to track the impact of charity research funding', 22 December 2014, http://www.amrc.org. uk/our-work/showing-impact/using-researchfish-track-impact-charity-research-funding (accessed 15 October 2015).

9. Research Councils UK, 'Research outcomes', 2015, http://www.rcuk.ac.uk/ research/researchoutcomes/ (accessed 15 October 2015).

10. Research Councils UK, 'Impact Report 2014', 2014, *Swindon: Research Councils UK*, http://www.rcuk.ac.uk/RCUK-prod/assets/documents/ publications/2014ImpactReport.pdf (accessed 15 October 2015).

11. Grant, J., and Wooding, S., 'In Search of the Holy Grail: Understanding research success', *Cambridge: RAND Europe*, 2014.

12. Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., and Johnson, B., 'The Metric Tide: Report of the Independent Review of the role of metrics in research assessment and management', *HEFCE*, 2014, http://www. hefce.ac.uk/media/HEFCE,2014/Content/Pubs/Independentresearch/2015/ The,Metric,Tide/2015_metric_tide.pdf (accessed 15 October 2015).

13. Sörlin, S., 'Funding diversity: Performance-based funding regimes as drivers of differentiation in higher education systems', *Higher Education Policy*, 2007, 20: 413–40.

14. ICNARC, 'CMP Participating Units', June 2015, https://www.icnarc.org/Our-Audit/Audits/Cmp/About/Participation (accessed 15 October 2015).

15. UK Clinical Research Collaboration, 'UK Health Research Analysis 2014', 20 August 2014, http://www.hrcsonline.net/pages/uk-health-research-analysis-2014 (accessed 15 October 2015)

16. Research Councils UK, 'Future Researchfish enhancements and interoperability', 2015, http://www.rcuk.ac.uk/research/researchoutcomes/future/ (accessed 15 October 2015).

17. McDowall, R.D., 'Fat Finger, Falsification, or Fraud?' *Spectroscopy Winter edition*, 1 December 2010, Vol 25, Issue 12.

18. Cai, L, and Zhu, Y., 'The Challenges of Data Quality and Data Quality Assessment in the Big Data Era', *Data Science Journal*, 2015, http://doi.org/10.5334/dsj-2015-002.

19. Wang, RY., and Strong, DM., 'Beyond Accuracy: What Data Quality Means to Data Consumers' *Journal of Management Information Systems*, 1996, 12(4) 5–33.

20. Magnus, KM., '*Data Management; Responsible Conduct of Research Education Committee (RCREC)*', September 2002, http://www.bu.edu/orc/files/2010/09/Session1Reading.pdf (accessed 15 October 2015).

21. Research Councils UK, 'Overview of Systems Interoperability Report', 17 June 2015, http://www.rcuk.ac.uk/RCUK-prod/assets/documents/publications/osipreport.pdf.

22. Ferguson, N., 'Researcher identifiers: National approaches to ORCID and ISNI implementation', *Knowledge Exchange Workshop*, June 22-23 2015.

Annexes

	Subcategories	Questions
1	**Publications**	
		Type of publication
		Chapter title
		Chapter author (surname and initials)
		Chapter pages
		Title (book, article, report, conference paper etc)
		Journal title
		Conference name
		First named author (surname and initials)
		Secondary author(s) (surname and initials)
		First named editor (surname and initials)
		Secondary editors (surname and initials)
		Volume
		Issue
		Pages (x-y)
		ISBN
		ISSN (print)
		ISSN (digital)
		ISSN (linking)
		Edition
		Is this published?
		Month of publication
		Year of publication
		Expected month of publication
		Expected year of publication
		Publisher
		Place of publication
		Available through open access
		Will the publication be available through open access?
		Publications URL
		Digital Object Identifier

2	**Collaboration and partnerships**	
		Please enter a short title for this collaboration or partnership for your reference purposes
		Please provide details of the collaborator(s) and/or partner(s)
		For each partner - please enter the organisation with which you have collaborated or partnered. This is a predictive input so you need only enter in the first few letters of the name. If the entry s not found, you will be able to enter it in full
		For each partner - Please enter the department with which you collaborated/partnered or leave blank if not applicable. Please do NOT enter individual names of people.
		For each partner - has this collaboration or partnership brought a direct financial contribution to your research?
		For each partner - please enter the currency of the contribution. This is a predictive lookup, so start typing the name of the currency and select the appropriate currency
		For each partner - enter the amount of any direct financial contribution made under this collaboration or partnership to the nearest unit
		For each partner - has this collaboration or partnership brought any in-kind contribution to your research?
		For each partner - Please enter the currency of the in-kind contribution. This is a predictive lookup, so start typing the name of the currenncy and select the appropriate currency
		For each partner - Please give an estimate of the value of the in-kind contributions made by your partners to this collaboration or partnership
		Briefly describe the contributions made by you and/or your research team to this collaboration or partnership
		Briefly describe the contributions made by your partners to this collaboration or partnership
		In which year did this collaboration or partnership commence?
		Is this collaboration or partnership still active? If not, in which year did it cease?
		If there is a URL which relates to this collaboration, enter it here
		List any outputs or otucomes that have resulted from this collaboration or partnership. Full details of each should be reported under the relevant sections of the form. Indicate whether this collaboration is multidisciplinary; if so, outline each of the disciplines involved

		Please categorise the impact of this collaboration or partnership using the check boxes below: cultural, societal, economic, policy and public services, no impact yet
		Is this collaboration or partnership governed by formal agreements such as material transfer agreements, or confidentiality agreements?
3	**Further funding**	
		Enter the title of the funding scheme for which you have successfully applied
		Enter the name of the organisation that provided the funding. (This is a predictive input, so you need only enter in the first few letters of the name. If the entry is not found, you will be able to enter it in full)
		Please select the most appropriate type for this funding
		Please enter the currency of the funding (eg GBP) (This is a predictive look-up: start typing the name of the currency and select the appropriate currency)
		Enter the value of the funding pledged/received from this scheme to the nearest whole currency unit. Please ensure you enter the value in full (eg 1 million is entered as 1000000)
		If known, please enter the grant/award reference number, otherwise leave blank
		Please enter the month in which the funding commenced/is due to commence
		Please enter the year in which the funding commenced/is due to commence
		Please enter the month in which the funding terminated/is due to terminate
		Please enter the year in which the funding terminated/is due to terminate
4	**Next destination and skills**	
		Please provide a label for the staff member you want to tell us about so that you can easily distinguish multiple responses. This is purely to help identify the data in the form and will NEVER be used in analysis or output from your organisations. We recommend that you use the staff member's initials
		Gender
		Please indicate what proportion of the full-time equivalent this team member had in your team (please enter a proportion of 1: for example, 0.5 would be half a full-time equivalent post)
		Role of the staff member when he/she was working in your research group
		Role of the staff member after he/she left your research group

		The sector into which the staff member moved
		Please select the industry sector/discipline to which the staff member moved
		The year in which the staff member left your team or the grant was terminated
		Select the country in which the staff member was subsequently based. (please enter unknown as appropriate).
		Select any qualifications gained by this member of staff in the course of working as part of your team (tick all that apply)
		Please describe other
		What month was this qualification gained?
		What year was this qualification gained?
5	**Engagement activities**	
		Please enter a short title for this activity for reference purposes (eg school visit (Lincoln), media interest (ageing study)
		What was the engagement activity? We are interested in the role played by you or a member of your team in the activity, regardless of whether this was presenting to or engaging with an audience
		Was this part of a recognised scheme, advisory panel, or was this part of a fundraising activity? Examples would include the National Advisory Committee on Drugs and the Edinburgh International Science Festival
		Please provide the details here
		Indicate the kind of presentation made
		What was the geographical reach of this activity?
		Who was the primary audience?
		In which year did this activity take place? For activities that have occurred every year or in multiple years, you can make multiple selections here
		What was the main purpose of this activity?
		How many people do you estimate this activity reached?
		Briefly describe the results of your activity. For example, 'talk sparked questions and discussion afterwards'
		Briefly describe any notable impacts that arose from this activity. For example, 'After my talk, the school asked if current GCSE students could visit my lab and also reported higher than expected interest in science GCSEs from pre-GCSE students'
		What do you consider to be the most important impact of this activity
		If there is a URL which relates to this activity, enter it here

6	**Influence on policy, practice, patients and the public**	
		Provide a short title or name for this influence on policy or practice. This is for reference within the form (eg citation in Cochrane Review)
		Select type/method of influence from this list
		Please enter title of guideline
		Please enter the publication which has been cited
		Please enter area of healthcare
		Enter name of organisation issuing guideline (This is based on a predictive look-up so you need only type the first few letters)
		Select the year in which the influence on policy or practice was first realised
		Select the option that best geographically represents the extent of this influence on policy or practice
		Enter the country (This question is based on predictive input: you need only enter the first few letters and then select from the options presented)
		Please select the area of policy influence - you can make multiple selections
		Describe other
		Has this influence on policy led to any of the following impacts?
		Briefly describe the impacts of this change in policy or practice. This should include (if applicable) the reach and significance of the impact, such as quantitative information regarding the benefits (increases in survival, quality of life, decreases in incidence, improvements in clinical service delivery, economic impacts, etc)
		If there is a URL which relates to this influence on policy, enter it here
7	**Research tools and methods**	
		Enter a short name for this research tool or method
		Select the type of research tool or method
		Briefly describe this research tool or method
		Have you provided this research tool or method to other research groups?
		In which year was the research tool or method first provided to other researchers?
		Briefly describe any notable impact(s) resulting from the development of this research tool or method

		If there is a URL which relates to this research tool or method, enter it here.
8	**Research databases and models**	
		Enter a short name for this research database or model. (This is for your reference purposes only)
		Select the type of research database or model
		Briefly describe this research database or model
		Have you provided this research database or model to other research groups?
		In which year was the research database or model first provided to other researchers?
		Briefly describe any notable impact(s) resulting from the development of this research database or model
		If there is a URL which relates to this research database or model, enter it here
9	**Intellectual property and licensing**	
		Select the phrase that best describes the protection this discovery/development has received
		Enter the patent application number (eg WO03075629). Please provide the earliest instance in whatever jurisdiction
		Name or title of the discovery/development
		Select the year in which this protection was received. You can select 'Unknown' if appropriate
		Briefly describe the discovery/development
		Has this intellectual property been formally licensed to others on a commercial or non-commercial basis?
		Briefly describe any notable impact(s) (other than licensing) that have arisen from this discovery/ development
		If there is a URL which relates to this intellectual property or licence, enter it here
10	**Medical products, interventions and clinical trials**	
		Enter the short name of the medical product or intervention for reference purposes
		Select the type of medical product or intervention
		Select the current active development stage for this medical product or intervention, or the stage most recently completed (as appropriate)
		Has the medical product or intervention been tested via a clinical trial? (If not applicable, please choose Not Applicable)

		Please enter the UKCRN/ISCTN reference ID here
		Select the year in which this development stage was completed
		Select the development status of this medical product or intervention
		Briefly describe the medical product or intervention, its current or most recent stage of development, and the current or most recent principal source of funding for this development
		Please choose the achievements that apply to your medical product intervention
		Please describe
		Briefly describe any other notable impacts arising from the development of this medical product or intervention. This can include impacts arising from the development process
		If there is a URL which relates to this medical product or intervention, enter it here
11	**Artistic and creative products**	
		Provide a short title or name for the artistic or creative product
		Select the type of artistic or creative product from the list
		Briefly describe the artistic or creative product
		Please select the year that this output was realised
		Briefly describe any notable impacts resulting from the development of this artistic or creative product
		If there is a URL which relates to this output, enter it here
12	**Software and technical products**	
		Provide a short title or name for the software or technical product
		Select the type of software or technical product from the list
		Was it Open Source?
		What Open Source licence is the software or technical product licensed under?
		Briefly describe the software or technical product
		Please select the year that this output was realised
		Briefly describe any notable impacts resulting from the development of this software or technical product
		If there is a URL which relates to this output, enter it here

13	**Spin-outs**	
		Enter name of the company
		Please enter the registration number of the company
		In which year was the company established?
		Enter the number of salaried people employed
		Briefly describe the company
		Briefly describe any notable impacts from this company
		If there is a URL which relates to this company, enter it here
14	**Awards and recognitiion**	
		Select the type of award or recognition
		Enter the name or title of the award or recognition (for reference purposes)
		Enter the name of the individual who received the award or recognition (only if this is already in the public domain, otherwise leave blank)
		Select the level of the award or recognition scheme
		Select the year in which the award was made or the recognition was received
		Briefly describe the award or recognition and the reason(s) it was made. Tell us here the role of the staff members who received the recognition.
		Briefly describe any notable impacts that have arisen from this award or recognition
		If there is a URL which relates to this award or recognition, enter it here
15	**Other outputs and knowledge**	
		Provide a short name/title for this output. (This will be used for your reference purposes only)
		Briefly provide details here, identifying any subsequent impact that may have arisen from your research
		Has this already been progressed further, either by you, by your collaborators or by other research teams?
		Please briefly provide details of this here. For example, what have you discovered?
		In what ways might this benefit society and/or the economy?
		To help us to continually improve the questions in the outcomes collection system, if you feel you should have been able to report this against one of the specific output types in the system, please select which one
16	**Use of facilities and resources**	
		Provide a short name/title for this work (this might be the name of the facility). (This will be used for your reference purposes only.)

	Please enter the name of the facility (This is a predictive input, so you need only enter in the first few letters of the name. If the entry is not found, you will be able to enter it in full)
	Briefly note the services provided or the nature of the resource shared through this (eg Beamline time, training, access to NMR machine, shared dataset, etc)
	Enter the year in which you began using this service/facility/ centre
	Enter the year in which you stopped (or expected to stop) using this service /facility/ centre
	Note any subsequent impacts from this work
	If there is a URL which relates to the work that is being reported, enter it here

Annexe B: Interviewees

Funders	Interview method
Science and Technology Facilities Council	Telephone
Tenovus Cancer Care	Telephone
Alberta Health	Telephone
Medical Research Council	In-person
Sparks Charity	In-person
Kidney Research UK	Telephone

Research organisations	Interview method
University of Leeds	Telephone
Imperial College London	In-person
University of Aberdeen	Telephone
University of Nottingham	In-person
Coventry University	In-person
University of Central Lancashire	Telephone

Annex C: Interview protocol

1. How long have you used the Researchfish system?
a. Funders only: is it mandatory for your grantholders to report back?

2. What Researchfish functions do you and your institution commonly use?

3. Do you use one of the following methods to collect additional data:
a. Department/institutional internal systems?
b. External systems?

4. Do you and your colleagues collect outputs from the Researchfish system? What outputs do you commonly collect?
a. How does your department/institution collect data on impact, and how is it measured?
b. Has Researchfish changed the way you report on the impact of collected data for your department/institution?
c. What do you do with the information, and how is it shared?

5. Please tell me about using Researchfish. What do you find to be the three main advantages of the system? What works?

6. What are the three most significant issues with Researchfish?

7. Looking to the future, what are the three possible opportunities for Researchfish?

8. What would be your 'utopia'?

9. Do you think integration with other external systems is important and/or necessary? If so, why?

10. [Funders only] How does integration of Researchfish with other systems work?

11. [Funders only] What has changed in your funding strategy since you started with Researchfish?

12. Can you tell us who governs the question set and how? Does that work for your institution?

13. [HEIs only] How do you report back to funders, and what type of information do you report? Why do you think they need this?

14. When faced with a technical issue do you use:
a. Internal IT support?
b. Researchfish support?

If you use both, which takes preference, and for what types of technical issues?

15. Does Researchfish meet your expectations?

www.kcl.ac.uk/sspp/policy-institute
@policyatkings