Provision and analysis of key indicators in research and innovation

Policy brief E – Testing the societal outcomes of R&I policies

Altmetric case study: Is cross-disciplinary research increasing the odds of research findings influencing decision-making?

Written by Etienne Vignola-Gagné, Henrique Pinheiro and David Campbell
March 2021

Research and Innovation

Provision and analysis of key indicators in research and innovation. Policy brief E – Testing the societal outcomes of R&I policies. Altmetric case study: Is cross-disciplinary research increasing the odds of research findings influencing decision-making?

European Commission
Directorate-General for Research and Innovation
Directorate G — Common Policy Centre
Unit G.1 — Common R&I Strategy & Foresight Service
Contact Athina Karvounaraki
Tiago Pereira
Alexandr Hobza
Email RTD-STATISTICS@ec.europa.eu
Athina.KARVOUNARAKI@ec.europa.eu
Tiago.PEREIRA1@ec.europa.eu
RTD-PUBLICATIONS@ec.europa.eu
European Commission
B-1049 Brussels

Manuscript completed in March 2021.

PDF	ISBN 978-92-76-41083-6	doi:10.2777/16075	KI-01-21-168-EN-N

Luxembourg: Publications Office of the European Union, 2021

EUROPEAN COMMISSION

Provision and analysis of key indicators in research and innovation

Policy brief E – Testing the societal outcomes of R&I policies. Altmetric case study: Is cross-disciplinary research increasing the odds of research findings influencing decision-making?

Prepared by

Science-Metrix

2021 Directorate-General for Research and Innovation EN

Table of Contents

Tables

Figures

EXECUTIVE SUMMARY

Science, technology and innovation policy has long been concerned with increasing the societal returns on its investments in experimental and developmental research projects. Policy interventions also often aim to address complex challenges (e.g. the United Nations Sustainable Development Goals), requiring input from a broad range of stakeholders.

One of the main policy instruments used to foster increased societal returns from research investments is to support interdisciplinary and multidisciplinary projects (henceforth, *cross-disciplinary research*). While definitions for these modalities of research vary greatly, here we define *interdisciplinarity* as the intellectual and experimental integration of concepts and tools from multiple scientific disciplines, whereas *multidisciplinarity* refers to the collaborative dimension of work accomplished by two or more scientists or scholars with different disciplinary backgrounds.

The assumption underlying this support for cross-disciplinary research is that it will fuel the desired policy-related returns. However, the argument linking cross-disciplinary research and societal outcomes is supported by only limited evidence of the ability of the first to bring about the second and whether policymakers succeed in fostering cross-disciplinary research practices. Multiple methodologies have been proposed to measure levels of cross-disciplinary research in scientific publications, and associated indicators are occasionally used in programme evaluations. But, in contrast, societal outcomes of research remain difficult to robustly measure at scale, as illustrated by the recourse to large-scale panels of peers in exercises such as the UK REF Impact Case Study component.

With the increasing policy attention on societal outcomes, the evaluation and bibliometrics community began to look for quantitative indicators that might capture societal outcomes of research (used interchangeably with *societal outcomes*, henceforth) at a large scale. Technometrics, especially when combined with bibliometrics, has so far been the main robust quantitative approach able to capture a restricted scope of societal outcomes.

In a recent development, the Overton database has applied automated parsing and matching tools to the realm of policymaking and governance archives. Launched by the founder of a pre-eminent altmetrics company (named Altmetric), Overton notably records mentions of peer-reviewed publications in the documents in these archives, enabling the measurement of 'citations' by the documents towards the scientific publications. The policy documents indexed by Overton include white papers, parliamentary records, and other executive and legislative archives, as well as knowledge reviews produced by governmental organisations providing scientific advice to policymaking. Overton also indexes documents produced by think tanks and intergovernmental organisations such as WHO or the OECD. The citations found in Overton records could therefore capture processes of knowledge transfer from academic research towards policymaking. If the robustness and validity of policy document citations can be confirmed, they would support a quantitative strategy to capture an invaluable new dimension when measuring societal outcomes.

With advances in text mining and web analytics, the field of altmetrics emerged around 2010 as a new and promising strategy for capturing societal outcomes. There are two understandings of *altmetrics*. The term can refer to any quantitative strategy aiming to capture societal outcomes of research at scale. Some now even place existing tools to capture societal outcomes, such as patent analysis, in that large umbrella category. The evaluation community also often refers to altmetrics specifically when discussing tools that track social media mentions towards peer-reviewed publications (those on Facebook, online journalistic news outlets, Wikipedia and Twitter, notably). Given the increased importance of demonstrating societal outcomes of research, expectations are that altmetrics could offer a crucial new strategy for the programme evaluation methodological toolbox. These expectations have yet to be fulfilled, however.

In this policy brief, these prior observations, problem spaces and emerging tools have been combined in a panel of given research questions, the first three of which will be directly answered here and the last two of which will be partially addressed:

A. Can altmetrics offer a practical, reliable, generalisable method for capturing societal outcomes of research in the form of informing policymaking?

B. Do higher degrees of cross-disciplinary research found in peer-reviewed publications increase the odds of these articles being cited in policy documents?

C. Can policy interventions (mainly funding) promoting cross-disciplinary research increase the odds of resulting findings supporting evidence-based policymaking?

D. Can we establish programme evaluation strategies that draw on findings from the validation of policy (mainly funding) mechanisms to enable pointed assessment of individual funding programmes or other policy instruments?

E. Can we establish a generic methodological framework to validate the societal outcomes of policy mechanisms for supporting research, using large-scale data sets? (That is, can our strategy to answer Question A be generalised to other cases of societal outcomes, other explanatory variables and other data sources?)

To answer these questions, a regression model was developed with Overton policy citation as the dependent variable (binary, cited or not) and the cross-disciplinary research level of FP7 and H2020 publications as the core independent variable. More specifically, two modalities of cross-disciplinary research were investigated: the disciplinary diversity of authors (DDA) and the disciplinary diversity of references (DDR). Additional tests were also performed to better characterise the content of the Overton database citing the FP7 and H2020 publications.

Descriptive features on the population of policy citations made to FP7 and H2020 papers confirm the likely feasibility of the strategy for evaluative bibliometrics. A manual investigation found a moderate number of false positives in a random sample of 50 policy citations (14 %). The comparatively large volume of observations of policy citations recorded in the population of publications considered, as well as their distribution (6 % of articles being cited once or more), indicates that these events are common enough to be expected in future assessments and to make their measurement worthwhile. The average time to policy citation peak (roughly 3 years after publication year) indicates that a policy-citation-based indicator could be deployed in most retrospective evaluation exercises.

Findings from the regression modelling of the association between policy citation and cross-disciplinary research (in addition to other dependent variables) show that DDA, but not DDR, appears to foster policy citations. On average, the probability of being cited in policy documents increases by roughly 1.2 percentage points after a typical increase (1 standard deviation) in DDA. This is a potentially relevant increase, considering the average uptake of 6 % for the entire sample.

At the level of FP7 and H2020 programmes, policy citation levels generally (but not always) correlated to recorded DDA levels and/or prior expectations given each programme’s goals. These findings indicate that DDA is not the sole research practice that fostered policy uptake of research findings, and that multiple research and engagement practices are combined in processes of evidence-based policymaking. In the absence of a fuller picture, however, policymakers and funding agency managers should generally add in provisions to support collaborative multidisciplinarity when designing funding programmes, if the fostering of policy-related outcomes of research are deemed an important component of these programmes.

Our large-scale findings of an association between cross-disciplinary research and policy citation can be used as supporting evidence in assessing the outcomes of specific funding programmes, where comparatively low volumes of publications will most likely prevent any definitive conclusions about how policy outcomes have been fostered. The approach followed in this brief could also be extended to other research strategies (altmetric or otherwise), in trying to develop a more comprehensive toolbox of quantitative methods to measure societal outcomes of research.

1 Introduction

As the emphasis funding organisations place on the longer-term socioeconomic impacts of research has increased, more funding programmes have promoted cross-disciplinary research, assuming it will fuel such policy-related returns (Gleed & Marchant, 2016). This argument linking cross-disciplinary research and societal outcomes is supported by limited evidence concerning the ability of the first to bring about the second (Chavarro, Tang, & Rafols, 2014) or whether policymakers can succeed in fostering cross-disciplinary research (Garner, Porter, Newman, & Crowl, 2012).

WHAT ARE ALTMETRICS AND WHAT IS AT STAKE?

There are two understandings of 'altmetrics'. The term can refer to any quantitative strategy aiming to capture societal outcomes of research at scale. Some now even place existing tools to capture societal outcomes, such as patent analysis, in that large umbrella category. The evaluation community often refers to altmetrics specifically when discussing tools that track social media mentions towards peer-reviewed publications.

Given the increased importance of demonstrating societal outcomes of research, expectations are that altmetrics could offer a crucial new strategy for the programme evaluation methodological toolbox. These expectations have yet to be borne out.

The literature review in this brief covered work on altmetrics writ large, including social media-based indicators. The experimental component of the brief is based on policy documents published by governmental and public agencies, not on social media analysis.

Altmetrics, in particular citations to the peer-reviewed literature in policy documents, may help to measure societal outcomes. However, attributing the longer-term socioeconomic impacts of research to specific funding programmes promoting cross-disciplinary approaches is difficult for several reasons:

(1) a lack of suitable data sets to confidently discard confounding factors (such as local and global trends in research systems, or the combination of impacts that comes with combining multiple streams of funding in research) in testing the effects of specific programmes using quantitative approaches such as econometric modelling and difference-in-differences (Buenstorf & Koenig, 2020; Hird & Pfotenhauer, 2017) – for example, data sets related to funding programmes specifically building on cross-disciplinary research are often too small to offer adequate statistical power in the complex model specifications required to resolve attribution;

(2) a lack of unbiased and well-characterised altmetric data sources, making it difficult to explore/rank the longer-term impacts of research; and

(3) uncertainty regarding the phenomena captured by altmetrics and their exact association with the goal of increasing societal outcomes from research (Haustein, 2016).

Recently, path-breaking studies have focused specifically on policy citations contained in the Altmetric.com database (note the distinction between the *Altmetric* brand and the *altmetrics* field of research). These studies reported that only low numbers of papers get cited in policy documents. They also reported a high skewness in citation distribution – with a high share of the overall citations going towards the applied life sciences and social science fields – and technical issues with the Altmetric.com database (Bornmann, Haunschild, & Marx, 2016; Haunschild & Bornmann, 2017; Tattersall & Carroll, 2018). Based on detailed content analysis of citing policy documents, Newson and colleagues found multiple instances of research mentions that were not made as formal citations, concluding that '[c]itation rates are likely to provide an underestimation of research use by policy agencies' (Newson, Rychetnik, King, Milat, & Bauman, 2018, p. 10). To address some of the shortcomings of prior studies measuring citations from policy documents towards peer-reviewed publications, as well as the obstacles facing the broader field of evaluative altmetrics, we made use of the Overton database, a new altmetrics resource dedicated to documenting some of the policy outcomes produced by research. Science-Metrix appraised whether peer-reviewed publications with a higher degree of cross-disciplinary research also saw higher propensity to be cited by policy documents, framing the research around the five questions presented in the text box below.

Statistical modelling was first performed on a sample of roughly 34 000 UK papers funded through the Framework Programmes (FPs) for Research and Technological Development (i.e. FP7 or H2020), and the extent of cross-disciplinarity was captured at the paper level through two lenses: the disciplinary diversity of contributing authors and the disciplinary diversity of cited references, which tracks diversity of integrated knowledge (see Section 4.2.1; see Section 3.2 for the corresponding methods). By first restraining the analyses to a sample of UK papers, we limited the possible effects of coverage biases in the Overton database (which is produced by a UK-based company and thus has a better coverage of policy documents from the Anglosphere). By selecting papers supported by FP7 or H2020, it was also possible to control for some project characteristics in assessing the association between cross-disciplinary research and subsequent policy uptake, in addition to increasing the relevance of the study's findings for the European Commission (EC).

The analyses were subsequently expanded to a larger pool of about 126 000 FP7-/H2020-funded papers from all European countries (see Section 4.2.2; see Section 3.2 for the corresponding methods). If the findings for the UK sample are confirmed on this larger data set, it will suggest that coverage bias may not be an issue in

The research presented in this policy brief was framed by the following five research questions:

A) Can altmetrics offer a practical, reliable, generalisable method for capturing societal outcomes of research in the form of informing policymaking?
B) Do higher degrees of cross-disciplinary research found in peer-reviewed publications increase the odds of these articles being cited in policy documents?
C) Can policy interventions (mainly funding) promoting cross-disciplinary research increase the odds of the resulting findings supporting evidence-based policymaking?
D) Can we establish programme evaluation strategies that draw on findings from the validation of policy (mainly funding) mechanisms to enable pointed assessment of individual funding programmes or other policy instruments?
E) Can we establish a generic methodological framework to validate the societal outcomes of policy mechanisms for supporting research, using large-scale data sets? (That is, can our strategy to answer Question A be generalised to other cases of societal outcomes, other explanatory variables and other data sources?)

It should be noted that the formulation of the first three questions was informed by an ambition to use them as the initial steps in building a more generic framework for the quantitative assessment of the societal outcomes of specific types of funding mechanisms, especially in the context of a common lack of suitable data on a programme basis. Research Questions D and E were then added. These high-level questions could not be fully addressed directly through the empirical evidence collected and presented here, although the current study aimed to achieve initial steps towards answering them.

similar studies covering non-Anglo-Saxon countries, acknowledging that not all European countries are equally represented in this broader data set. Indeed, the larger players (e.g. France and Germany, to name just a few) weigh heavily here and are still better covered in Overton than the smaller Eastern European countries. It will also make the study's conclusions more generally applicable to the broader European context. Finally, descriptive statistics for the key models' variables were disaggregated by discipline and sub-programme to highlight key differences across the FP7 and H2020 funding mechanisms (Section 4.3).

Before moving on to the description of and results from the empirical investigation (Section 3 and Section 4, respectively), the following pages will first review the currently available evidence on a chain of prior assumptions (as listed below and widely circulating in the academic and policy communities) that have led to the formulation of the core research questions explored:

(1) science, technology and innovation (STI) policy increasingly supports cross-disciplinarity research (Section 2.1),

(2) cross-disciplinary research practices can be fostered through funding instruments and other policy interventions (Section 2.2),

(3) cross-disciplinary research leads to improved societal outcomes (Section 2.3),

(4) altmetrics could offer a robust quantitative strategy to capture societal outcomes from research (Sections 2.4 to 2.6), and

(5) altmetrics relying on policy documents citing the scientific literature could offer a robust quantitative strategy to capture societal outcomes from research specifically on the topic of (governmental, NGO or think tank) decision-making (Section 2.7).

These assumptions, as explained above, do not directly map onto the research questions explored here. Rather, they synthesise prior knowledge as well as the known gaps in knowledge that enabled Science-Metrix to initially prepare the experimental design deployed for this study.

2 Literature review

2.1 Policy interest in cross-disciplinarity

Policy interventions often aim to address complex challenges (e.g. the United Nations Sustainable Development Goals (SDGs)), requiring input from a broad range of stakeholders. The diversity of stakeholders needed to inform such interventions can span multiple dimensions, such as their activity sector, geographic location and disciplinary background. With the increasing emphasis funding organisations place on the longer-term socioeconomic impacts from research, an increasing number of funding programmes promote scientific collaboration across these dimensions, assuming it will fuel such returns. The following is a small sample of pre-eminent policies and interventions targeting interdisciplinarity and boundary-spanning collaboration as policy goals:

- the EC FP's Responsible Research & Innovation agenda
- the EC FP's European Research Council (ERC) funding mechanism
- H2020 and especially its Future and Emerging Technologies and Societal Challenges pillars (LERU, 2016)

- EC COST actions, including thematic subfunding programmes such as BiodivERsA, which emphasise interdisciplinarity

- multiple US National Science Foundation (NSF) initiatives, including Convergence awards, Research Coordination Networks, and Science and Technology Centers

- large areas of research at the US National Institutes of Health (NIH) that fall under the concept of the 'translational sciences'

- a small group of multilateral or intergovernmental funders including, notably, the Human Frontier Science Program (HFSP), the Belmont Forum and Future Earth

- a set of interventions that fall within the category of 'excellence' competitive national funding to universities, including the Canada First Research Excellence Fund (CFREF), China's Double First Class University Plan and its predecessors, Germany's Exzellenzinitiativ, and Japan's Word Premier International Research Center Initiative

An international survey of national research funding agencies sponsored by the Global Research Council – a multilateral knowledge exchange mechanism for more than 20 national funding councils – found that, although '[m]ost of the funding agencies we interviewed were open in stating that they do not have formal policies relating to interdisciplinarity, [they] do have practices to encourage and support it' (Gleed & Marchant, 2016, p. 8). Support for cross-disciplinary research can therefore be safely considered a ubiquitous feature of STI policy in 2020.

But what justifies this flurry of policy interventions? Interdisciplinarity is advocated as the preferred tool to realise a number of central policy objectives for governments and societies. As a member of an EC expert committee on Research, Innovation and Science Policy put it, fostering interdisciplinary research could result in

> crossing departmental boundaries and inter-disciplinarity to generate new knowledge of transformative power… exploit[ing] new types of problem-driven and user-oriented R&D research programmes that go way beyond well-established modes of targeted, incentivized R&D top-down… Stimulat[ing] disruptive innovations to accelerate value creation across different industries and branches of knowledge through intellectual fusion, combinations and interfaces (Allmendinger, 2015, p. 4).

Funders and policymakers aim to support intersectoral collaborations just as much as they do academic collaborations across disciplines, notably with a view to intensifying knowledge transfer to boost the socioeconomic returns from research. Academia–industry collaboration, public engagement, co-production and other related concepts are all loosely associated with cross-disciplinarity. Again, this set of interventions is justified through highly ambitious projections of future benefits:

> co-production is understood as a way to enhance scientific accountability to society ('logic of accountability'), to ensure the implementation of scientific knowledge in society ('logic of impact'), and to include the knowledge, perspectives and experiences of extra-scientific actors in scientific knowledge production ('logic of humility') (van der Hel, 2016, p. 165).

2.2 Fostering of interdisciplinary research practices through funding instruments and other policy interventions

Of the assumptions that underpin the research presented here, perhaps the most fragile is the one that policy interventions can foster increased interdisciplinarity in the research groups they target. For instance, many studies documented trends towards increased interdisciplinarity in research, but without specifically linking this shift to policy interventions (Dworkin, Shinohara, & Bassett, 2019; Okamura, 2019; Porter & Rafols, 2009); or, as we just did above, they note the multiplication of interdisciplinarity initiatives and narratives originating from policymakers. If proved effective, these initiatives could be very important in fostering cross-disciplinary research as there is quantitative evidence demonstrating that traditional grant mechanisms tend to be conservative and to shy away from cross-disciplinarity. Bromham and colleagues, examining the interdisciplinarity and multidisciplinarity intensity of proposals to the Australia Discovery grants, found that interdisciplinarity in proposals was 'consistently negatively correlated with funding success.' Multidisciplinarity was positively correlated with peer-review scores but at a very small magnitude (Bromham, Dinnage, & Hua, 2016).

Most of the restricted body of work on the policy mechanisms through which funding interventions foster interdisciplinary research is qualitative and based on case studies, often resulting in recommendations for the management of these programmes (Lyall, Bruce, Marsden, & Meagher, 2013; Molas-Gallart, Rafols, & Tang, 2014). Elsewhere, programme evaluations have used peer-review panels to assign scores to projects or initiatives and therefore measure achievements in interdisciplinarity in a semi-quantitative manner (European Research Council, 2018).

Lyall and colleagues find that despite high-profile initiatives and policy exhortations to engage in interdisciplinarity, transdisciplinarity and/or knowledge transfer, still only a modest volume of STI policy practices in the UK meaningfully engage with these approaches in practice (Lyall, Meagher, & Bruce, 2015).

Porter, Garner and Crowl (2012) have provided one of a few specific evaluations of a policy instrument's effect on levels of interdisciplinary integration within supported scientific projects. The authors characterised the set of publications originating from the US NSF Research Coordination Network. This programme aimed to foster novel research networks around interdisciplinary intellectual projects. They found this programme to have succeeded in achieving high networking and interdisciplinarity metrics in related papers, although the authors noted that successful applicants to the programme already displayed higher scores on these dimensions before the support period in comparison to non-successful applicants. Similarly, Science-Metrix, using a difference-in-differences approach (the control group was selected using a regression discontinuity design), quantitatively demonstrated a positive association of one of HFSP's funding mechanisms (i.e. cross-disciplinary fellowships) on the level of interdisciplinarity achieved by its awardees (Science-Metrix, 2018). While both the awardees and control group scored highly prior to funding, HFSP funding appeared to have enabled the former group to maintain its level of interdisciplinarity during funding, whereas this was not the case for the latter. A sustainable and positive effect was also perceptible after funding for awardees who did increase their score by a greater margin than the control group by that time. However, the authors noted the lower reliability of the findings for this group given its size, and most of the other HFSP funding mechanisms did not appear to further increase the level of interdisciplinarity of the awardees. Still, HFSP stood out well relative to other funders for the overall interdisciplinary level of its supported papers. Science-Metrix is currently performing multiple evaluations of interdisciplinary funding (e.g. for the Belmont Forum and for CFREF).

2.3 Prior evidence on improved societal impacts for cross-disciplinary research

By the early 2010s and onwards, there appeared to be 'a consensus in the literature that socially relevant research is most often interdisciplinary' (Chavarro et al., 2014). Despite this consensus, broad-scope quantitative evidence on the capacity of cross-disciplinarity research to produce improved societal outcomes was not forthcoming. Interdisciplinarity may have become somewhat conflated with the notion of intersectoral collaboration or engagement, which is a pre-condition of knowledge and technology transfer, the latter itself being a clear instance of societal impact. The vast literature on technology transfer, academic entrepreneurship and knowledge societies may have underpinned the emergence of a consensus on the societal relevance of interdisciplinarity. Still, there is surprisingly little in the way of overt, generalisable evidence to support this collective assumption, especially in a way that applies to multiple pathways and modalities of interdisciplinary practice.

To review the literature on the contributions of academic entrepreneurship and mode 2 research practices to societal outcomes would be out of scope in the current brief.(Bozeman, Fay, & Slade, 2012; Etzkowitz & Leydesdorff, 2000). The research that has focused on a stricter definition of interdisciplinarity has, for its part, mostly relied on case studies. Disciplinary diversity in researchers' background has been found to be associated with increased chances to engage in entrepreneurship and technology transfer (Deste, Mahdi, Neely, & Rentocchini, 2012). Qualitative research has also shown that stakeholders and users in interdisciplinary projects share the perception that the approach is conducive to generating useful outcomes for these stakeholders' problems, although the extent to which these perceptions were realised was highly dependent on the type of strategies used (Molas-Gallart et al., 2014).

On the quantitative side, Chavarro and colleagues found that, in a set of WoS publication records with at least one co-author from Columbia, papers with higher scores on certain (but not all) dimensions of interdisciplinarity, were associated with a greater orientation towards local issues (Chavarro et al., 2014). Campbell and colleagues also found that the odds of research uptake in the patent literature was positively and significantly related to the multidisciplinarity of research teams. (Campbell, Struck, Tippett, & Roberge, 2017). Wang and Li found similar results looking at the effect of the scope of integrated knowledge on uptake in patents (L. Wang & Li, 2018).

2.4 Altmetrics to measure societal outcomes of research – an overview

In the decade spanning 2010–2020, a novel quantitative research evaluation tool emerged with the launch of databases recording the uptake of journal-based (or proceedings-based) scientific outputs beyond the scientific literature in, for example, social media, blogs, news and educational resources. These data, because they track usage beyond academic circles (although this is not always true) as traditionally captured in bibliometric indicators, are often referred to as alternative data (or altmetrics). Included in the databases' coverage are platforms such Facebook and Twitter, a selection of blogging platforms, journalistic and news websites, Wikipedia, Reddit, Stack Exchange and library holding databases. These mentions are usually tracked through document identifiers such as DOI, PMID and the URL of the article. The bibliometric scientific community has invested much attention and effort in the development of these indicators in the hope that they may provide improved measurements of the societal impact of science.

The value of social media mentions to journal articles is that they may capture degrees of readership, uptake and engagement, in an audience that is theoretically not restricted to peers. Expectations for the contributions of altmetrics to decision-making and evaluation have been high, as illustrated by the contentions of an expert group on altmetrics recently convened by the European Commission:

> Altmetrics also have potential in the assessment of interdisciplinary research and the impact of scientific results on the society as a whole, as they include the views of all stakeholders and not only other scholars (as with citations). Hence, altmetrics can do a better job at acknowledging diversity (of research products, reflections of impact etc.), providing a holistic view of users as well as providers of scientific products, and enhancing exploration of research results (European Commission Expert Group of Altmetrics, 2017, p. 11).

Further, the same group summarises the potential advantages of altmetrics as broadness (inclusion of multiple stakeholder types), diversity (type of outputs measured), multifaceted (different signals for a given output), and speed (readership of an article typically taking place faster than the uptake of its findings in ulterior research).

A US National Information Standards Organization (NISO) report identifies three main 'use cases' for altmetrics: supporting discovery (including providing analytics for strategic planning), research evaluation (with a specific focus on societal impacts, and including the demonstration of return on investment as well as benchmarking), and the showcasing of achievements (towards more distant stakeholders including the taxpaying public or politicians) (National Information Standards Organization, 2016). The report also provides a typology of potential users of altmetrics evidence that includes librarians, research administrators, funders or academic hiring committees, to name just some examples. By and large, these use cases are very similar to what can be achieved with bibliometrics, although with a focus on societal outcomes rather than research excellence.

The most common sources of observations on altmetric uptake are as follows:

- Facebook
- Twitter
- Wikipedia
- Mendeley
- Reddit
- F1000Prime
- CrossRef
- CiteULike
- multiple blogging platforms, such as Nature Blogs, Wordpress.com and others

Additionally, some altmetrics service providers also parse and aggregate – within coherent data sets –dispersed observations drawn from activity sets such as the following:

- governmental-, intergovernmental-, NGO- and think tank- produced policy documents, and
- journalistic news outlets, including online magazines.

Ortega's review of altmetrics service providers include the following main players:

- Mendeley, a service that tracks readership of journal and proceedings publications on its own reference manager and reading platform but that simultaneously acts as a major service provider
- Lagotto, developed by the Public Library of Science (PLoS) in 2009 as the first aggregator of data from multiple altmetric sources for publications from journals edited by a small group of publishers
- Altmetric.com, arguably the pre-eminent aggregator of altmetric data streams and certainly the service with which or on which the highest volume of research has been conducted
- ImpactStory, an aggregator focusing on researcher-level personal profiles (and one that does not offer access for at-scale production in either a research or evaluation context)
- PlumX, an aggregator that includes many common sources as well as article usage metrics (views and downloads)
- Crossref Event Data (CED), a non-commercial provider providing raw data on altmetric interactions (one step earlier in the processing chain than commercial aggregators) (Ortega, 2020)

From this list, Lagotto and ImpactStory have seen only limited deployment. Mendeley and Altmetric.com are the most studied sources, with PlumX recently emerging as an alternative broad-scale aggregator to Altmetric.com. Ortego reported that the CED service may prove useful in improving coverage of Wikipedia mentions.

New altmetrics databases are still emerging, as is the case of the Overton policy database that will be deployed in the empirical component of this study. Arguably, other analytical strategies such as examining citations to scientific publications from patent or clinical guideline records can also be included within the broader definition of altmetrics, especially as the field relates to broad societal outcomes of research (Tahamtan & Bornmann, 2020).

Like citation counts, observations on altmetric mentions or interactions can be processed in multiple ways to compute different indicators. Again as in computations with citation data, altmetric observations have been shown to be shaped by varying disciplinary features, temporal trends and database coverage biases, meaning that raw volume counts are almost never useful (Thelwall, 2016). Basic normalisation procedures used for citation indicators can also be applied for altmetric indicators (normalisation by subfield and by year). Thoughtful interpretation of altmetrics findings should also consider the limitations presented in the following section.

To conclude this introductory overview of altmetrics approaches, we note that most of these strategies are geared towards the capture of broader societal attention towards scientific publications issued in journals or conference proceedings. Yet it could be argued that non-scientific or hybrid outcomes are increasingly becoming the focus of transdisciplinary, co-productive or locally oriented research projects (Koier & Horlings, 2015). Altmetrics approaches have still to be convincingly deployed for these kinds of outcomes, and the collective amount of explorative efforts conducted to try and do so has been low. Assessment of these kinds of outcomes must – until large-scale efforts for their indexation materialise – make use of qualitative, expert review or survey methods.

2.5 Challenges in the implementation of altmetrics

There is strong demand for robust quantitative methodologies from the evaluation community, driving a steady flux of developments and proposals on the evaluative use of altmetrics. It must be heavily emphasised that altmetrics research findings have yet to be systematically transposed to the programme evaluation context (see the following section for an examination of the few relevant examples available). At this stage of the field's development, in fact, most experimental efforts remain oriented towards surveying the basic features and behaviours of online audiences towards scientific publications, as well as the basic features of the databases that record these audiences' practices.

Issues of distribution and prevalence of altmetric uptake remain central to current research on altmetrics (Haustein, 2016; Thelwall, 2018). It is still unclear whether prevalence of altmetric mentions is low or high in a given subfield, and what levels can be considered as reference points. There are also many technical issues with the use of the various altmetric database aggregators that are still being identified and addressed, including discrepancies in results obtained by these various sources for a fixed publication set (Zahedi & Costas, 2018).

Perhaps the main obstacle currently facing the development of altmetrics methods for evaluation is the uncertainty about their capacity to capture meaningful instances of socioeconomic impact of research. Summarising findings from multiple studies, many have concluded that there are still many unknowns regarding the motivations for altmetric citations to peer-reviewed articles (Haustein, 2016; Tahamtan & Bornmann, 2020). Thelwall and Kousha have argued that only clinical guideline citations and patent citations would unambiguously capture societal outcomes of research outcomes (Thelwall & Kousha, 2015). Pulido and colleagues conducted in-depth examinations of the content of Twitter and Facebook posts on scientific articles, with the aim of determining whether these posts provided evidence of societal change achieved through the research (rather than online discussion and interaction strictly, with social change defined as 'the actual improvements resulting from the use of this knowledge in relation to the goals motivating the research project (such as the United Nations Sustainable Development Goals)') (Pulido, Redondo-Sama, Sordé-Martí, & Flecha, 2018). They found that this was only the case in 0.5 % of social media mentions to more than 5 000 journal articles from EU-funded projects.

Altmetrics findings often include observations from multiple sources (e.g. Facebook, Twitter and Wikipedia), but this may conflate observations on dimensions that are in fact very different from one another. Even considering only Twitter observations, one might question whether original tweets should be counted with or separately from re-tweets (Wouters, Zahedi, & Costas, 2018).

Science-Metrix has also been able to make a few observations on the limitations of altmetrics as part of its research evaluation projects. The value of these mentions, given how general a 'mention' is as a category, is hard to interpret meaningfully on its own. It might be argued that the audience for the discussion of scientific findings on social media is made up of scientists, rather than or as much as of the lay public. Additionally, it should be kept in mind that members of a research team may themselves refer to their own research on their social media pages. In this case, altmetric 'citations' are more representative of self-promotion than broad societal uptake. Disaggregating altmetric citations by source and using different metrics within each source may help to distinguish between cases of self-promotion and uptake. Mentions in news outlets or on Wikipedia can more safely be assumed to amount to broad uptake, for instance. Similarly, indicators based on highly mentioned publications may be less sensitive to self-promotion, compared with indicators based on the share of publications mentioned in a specific altmetric source, except perhaps for publications with many authors such as those resulting from research consortia. Accordingly, additional normalisation by number of authors might be desirable to properly control the effect of self-promotion. Further qualitative research would be needed to inform such decisions.

Another source of uncertainty concerning the capacity of altmetrics to capture socioeconomic outcomes is the novelty of the social media or other phenomena they capture. A general problem of capturing socioeconomic outcomes of research is that they may well take as many as 10, 15 or 20 years to fully materialise (Tahamtan & Bornmann, 2020). Programme evaluations generally tend to take place too early for such periods of realisation to have elapsed. Even where retrospective assessments of societal outcomes originating from supported projects are conducted by funders using expert panels, or qualitative case study methods, these exercises still tend to capture potential rather than realised impacts (Langfeldt & Scordato, 2015). Given this, it is doubtful that altmetric impact measurements made comparatively early after the publication of an article can capture the broad set of impacts deemed most desirable by policymakers and other stakeholders.

2.6 Current evaluation practices with altmetrics

Very few examples of formal programme evaluations that have included an altmetrics component were publicly available and online as of April 2020.

An altmetrics assessment of ERC-supported outputs, conducted by Rand Europe and the Montreal-based Observatoire des sciences et des technologies, provides a lone exception (Larivière et al., 2015). The evaluation was conducted at an early stage in the development of altmetrics and included Altmetric.com data for ERC papers published between 2012 and 2013, and Mendeley data for the period 2009–2013. Comparators were carefully selected from unsuccessful ERC applicants, EC FP, NSF, NIH and other groups of awardees of prestigious research funding. Comparisons were also broken down for senior and junior researchers, as well as by broad scientific domains (normalisation had not been achieved in this exercise). The exact findings obtained are of less interest here than the methods used, so suffice to say that ERC awardees clearly performed better on altmetrics dimensions than non-successful applicants did, and that their performances vis-à-vis comparator groups were roughly similar.

Koier and Horlings roughly assessed the potential of altmetrics for the evaluation of a Dutch transdisciplinary research programme with a high orientation towards local relevance and knowledge transfer (Koier & Horlings, 2015). This programme resulted in a high volume of outputs and outcomes that were not scientific publications, but for which online attention and societal uptake measures would have been desirable. These authors noted that at the time of their evaluation, altmetrics were yet to provide any viable pathway for the measurement of altmetric uptake for non-journal outputs. Their own attempts to use download statistics from the programme's centralised online repository had not yielded meaningful findings, and they noted the lack of a standardised, centralised database of records on non-journal outputs as a major impediment in the conduct of such analyses.

Guidelines for research evaluation using altmetric indicators may be more common than actual examples of such evaluations (Haustein, Bowman, & Costas, 2016). Notably, Wouters and colleagues provide in their design principles for altmetric evaluations examples of findings for altmetrics counts, mean counts and coverage measurements for Africa, the EU28 and the United States (Wouters et al., 2018). They also present examples of thematic clustering analyses and geographical mappings of EU28 and African publications with colour gradation according to coverage measurements for each of these subfields. They also represented Twitter communities of attention with network analysis, using shared hashtags as edges between unique users (nodes). The authors conclude that altmetrics focusing primarily on social media sources may not be fully soluble within existing evaluation practices given marked discrepancies in the objects of interest and standards of validity; rather, they call for new and different evaluation designs to be jointly developed with the new altmetric approaches.

Based on the above review of the literature, it is Science-Metrix's view that altmetrics might currently be best used in investigative rather than summative components of programme evaluations – that is, to answer narrow and well-bounded research questions to provide ancillary but not core evidence to answer central evaluation questions. That is, altmetrics can be used to provide in-depth answers to narrow questions, but the costs and resources required to produce robust findings most likely prevent their use at scale – for example, in dashboards and scoreboards built for tens or hundreds of groups, institutions and/or countries. This is the approach later deployed in the experimental section of this policy brief. Of course, this situation is expected to evolve rapidly given the highly dynamic nature of research on altmetrics.

2.7 Policy documents as an altmetric source and measuring the outcomes of research on decision-making

Preliminary work from a handful of articles using the Altmetric.com database makes it possible to infer some of the basic features of policy documents as a source of altmetric information. To our knowledge, no work has been produced yet on other altmetric databases covering policy documents, such as Overton. We are also not aware of a publicly available formal programme evaluation that has made use of policy document citations (Science-Metrix is currently running such an evaluation for the Belmont Forum).

Haunschild and Bornmann (2017) have examined policy document citations from the Altmetric.com database to a publication set consisting of more than 11.25 million WoS indexed articles issued between 2000 and 2014. They found 0.32 % to have at least one policy citation. The set of papers for the year 2005 displayed the highest share of policy citation (almost 0.5 %), indicating potentially much longer lags from publication to citation peak year in comparison to citations from other journal articles. Publication sets in the fields of Agricultural Economics and Policy (2.97 %), Tropical Medicine (2.64 %) and Economics (2.18 %) had the highest chances of receiving policy citations.

Bornmann, Haunschild and Marx (2016) examined policy citations from the Altmetric.com database towards records in their custom-built set of more than 190 000 papers on climate change. A share of 1.2 % of these papers had at least one policy citation in altmetrics. Of these papers, 78.7 % received only one policy citation. The authors found citation peaks to occur between two to four years after publication, but those documents with the highest levels of policy citations had citation peaks occurring later than the overall figure.

Tattersall and Carroll (2018) considered policy citations to journal articles published by the University of Sheffield. They report a share of 1.41 % of the overall Sheffield publication set to have been cited by at least one policy document. The disciplinary distribution of these citations very much followed what has been reported above for other policy citation studies, and studies that capture other altmetric dimensions. Much like in bibliometrics generally, citation distributions were also highly skewed with only a few articles achieving citation counts above 1. One finding from this team is worrisome: manually validating 21 policy citations to University of Sheffield articles, they found 7 for which attribution to the University of Sheffield or to the Sheffield article was problematic. Another finding from this study that acts as a call for caution is that there were a number of duplicate policy citations in the Sheffield set, sometimes because individual chapters of a full policy report are published separately. Additionally, some of the policy citations were found to originate in journal articles rather than actual government reports.

Newson and colleagues have used a 'backward tracing' approach to understanding policy citations, starting from policy documents and trying to characterise how they use citations. They selected a number of Australian policy documents relating to the topic of childhood obesity. These 86 childhood obesity policy documents made a total of 526 unique references to topically relevant research content, of which half were peer-reviewed publications and a fifth were non-peer-reviewed research publications. They concluded that in many cases (they did not compute a share of the overall citation data set), textual context for the citations does not make it possible to unambiguously attribute impact on the policy process for the research findings cited. As in citations within the scientific community, purposes and intentions for making a citation appeared diverse. The authors also found multiple instances of mentions to research that were not accompanied by an attendant formal citation, concluding that '[c]itation rates are likely to provide an underestimation of research use by policy agencies and the method has the potential to miss research that was in fact impactful, and place undue importance on cited research' (Newson et al., 2018, p. 10).

As previously mentioned, in recent years some researchers have turned to the mapping of altmetric relationships to go beyond the simple count of mentions and reposts:

> We suggest that the potential of social media data for impact assessment lies in that it may help understand the social interactions of researchers, both within and outside the scientific sphere. This implies a shift from the focus on the quantified 'impact' of current altmetrics indicators towards the use of social media networks to reveal the diverse contexts in which researchers potentially participate (Robinson-Garcia, van Leeuwen, & Ràfols, 2018).

Robinson-Garcia and colleagues applied this general strategy to the Twitter profiles of two colleagues. They manually coded their Twitter interactions and found out that for one researcher the Twitter interaction network was representative of a number of engagements with the policy world as well as with local actors. Given the amount of manual work involved in this project, it is doubtful, however, that such an approach can be deployed at scale.

Noyons and Ràfols have used Altmetric.com database policy citations in overlay maps of the structure of scientific fields (Noyons & Ràfols, 2018). Focusing on agricultural research, they found certain subtopics to be more likely to be associated with policy engagement than others, mostly subtopics in the social and behavioural sciences, health sciences, soil sciences and climate sciences.

3 Methods

3.1 Experimental design

Despite the shortcomings generally identified for altmetrics approaches, policy citations stand as the next best candidate to sit alongside patent citations and clinical guideline citations within the upper tier of comparatively reliable altmetrics indicators (Wilsdon et al., 2015). Policy mentions stand a high chance of capturing a well-defined societal impact, in the form of a scientific contribution to evidence-based policymaking (Bornmann et al., 2016). Citation in policy documents was one of eight indicators rated as highly important for the evaluation of societal outcomes by the stakeholders consulted by Willis and colleagues (2017).

Prior studies on policy citations used the Altmetric.com database. Here, we made use of Overton, a novel database established with an explicit goal to increase the coverage and comprehensiveness of policy-focused altmetrics. Overton records are built from combining a broad panel of government sources with web crawling. The base list of governmental sources in Overton includes a long tail of repositories with just a few documents each. The database indexes more than 2 million policy documents produced by governmental entities, international governmental organisations and think tanks. While close to 75 % of these records are provided by US, UK and intergovernmental sources, the database also contains more than 100 000 entries from Japan and 70 000 from Germany, to take just some examples. Overton coverage extends to the year 2019.

Using the Overton database here enables us to investigate research Question A. By investigating both cross-disciplinary research and policy uptake in a large set of publications (FP7 and H2020-supported publications), we are able to provide a robust answer to Question B. If both Question A and Question B are positive, it logically follows that Question C is also positive. Questions D and E cannot be fully answered by the experiment presented in this brief, but recommendations for future investigations can be formulated based on the partial evidence that was collected here on these dimensions.

3.2 Empirical deployment

3.2.1. DATA PREPARATION

Lists of publications produced through FP7- and H2020-supported projects were obtained from OpenAIRE and CORDIS and matched to Scopus and Overton. Choosing papers that could be matched to specific grants made it possible to control for some funding characteristics (see below). The outcome variable was coded as 1 for papers cited at least once in Overton records and 0 otherwise.

The data set was restricted to publications produced until 2016 (inclusively) allowing, at a minimum, for a four-year policy citation window (publication year plus three). This choice balanced the need to maximise the number of observations with information on the lag from publication to policy uptake (see Figure 1 in Section 4). Statistical models were tested with and without the publication year as a control, knowing that older papers have a higher chance of having been cited in a policy document. The final data set contained 126 441 papers published between 2008 and 2016. These papers were paired to FP7/H2020 project funding, sometimes more than one, resulting in ~137 000 observations in total.

Cross-disciplinarity at the paper level was captured through two lenses: disciplinary diversity of cited references (DDR; tracks diversity of integrated knowledge) and disciplinary diversity of contributing authors (DDA). The former is equivalent to the integration metrics of Porter and Rafols (2009) and relies on Science-Metrix's journal-based classification of science to classify a paper's cited references by subfield[1]. The

[1] See https://science-metrix.com/?q=en/classification

latter measured diversity as reflected in the prior disciplinary background of a paper's co-authors (team multidisciplinarity). Authors were disambiguated using Scopus author IDs, which produce reliable results at scale, when producing indicators on an aggregate of 1 000 or preferably 2 000 authors or more (Campbell & Struck, 2019). Science-Metrix subfields were assigned to authors based on their prior publications. A paper's DDA was computed by adapting the metrics by Porter & Rafols to the disciplinary profile of co-authors. DDA was designed to increase for teams involving authors from different subfields, particularly where these subfields are not frequently connected in Scopus. DDA and DDR were normalised by subfield to avoid coverage biases (Campbell et al., 2015). Other diversity indicators were computed and used in the modelling work, including the share of women authors, number of authors and number of countries. Findings for these three control variables are included in Table 1 for reference, but their coefficients were not interpreted in discussing the main conclusions presented below. Their inclusion in the study models was simply intended to control for confounders that could have biased the interpretation of the coefficients for the core study variables.

The model's specification accounted for additional characteristics of a paper: subfield/year-normalised (scientific) citation counts and CiteScore, document type, average number of prior papers per author. It also accounted for specific characteristics of the research projects that were fixed for the papers published under a specific research project during the period of analysis (e.g. researcher proximity to policymaking, amount funded, main topic of interest).

3.2.2. MODEL SPECIFICATION

The goal of the following quantitative exercise was to provide evidence to directly test research Question B, on the link between cross-disciplinary research and policy documents. Policy uptake was measured as a binary variable. This justified the choice of logistic regression to test the associations between bibliometric variables and policy uptake. The estimated coefficients linked the explanatory variables to the odds that a scientific publication impacts policy. The logit function is expressed as

$$log\left(\frac{p}{1-p}\right) = \alpha_k + \sum_{1}^{q} \beta_j x_j + u$$

Conditional logistic regression was used because papers from the same project may be more similar (Agresti, 2012). Roughly, this specification allowed for each research project to have a different baseline impact on policy. In the equation above, this is represented by the subscript *k* in the *α* intercept, which was allowed to be different for each research project.

This model is more suited to accounting for the unobserved characteristics of different projects that could affect estimates in 'usual' logistic regressions (if we were able to measure them). For example, the chances of being cited in written policy may be higher if the lead of a research team actively collaborates with policymakers.

The modelling strategy includes three models to account for different ways to measure the explanatory variables. The first model (log-transformed) used the logarithmic form of highly skewed explanatory variables (normalised citation counts, normalised CiteScore, DDA, DDR, number of authors/countries, and average number of papers per author) to reduce the effect of outliers. The remaining, less skewed variables were not transformed. The second model was based on the original form of all variables. In the third model the explanatory variables were divided by their standard deviations (except for the variable 'review'). Therefore, the odds ratio from this model refer to the changes in the odds of policy uptake associated with one standard deviation in the explanatory variable. This allowed for comparisons across variables in Model 3, although caution is advised since changes of one standard deviation in highly skewed variables are usually less likely. The standardised (third) model was the base of the main observations regarding effect sizes in this study. Model 3 was then tested in different subsets of the main data set in Table 2 to check whether the results observed for the UK-authored papers are also observed in non-UK papers and in the set containing all papers.

MODELLING PARAMETERS:

- Three different models to test the link between cross-disciplinarity and policy uptake.
    - In Table 1, Models 1 to 3 are based on the subsample of UK-authors and account for different ways to measure the explanatory variables (see below).
    - In Table 2, Model 3 (based on UK papers) is compared with sets of non-UK papers and all papers in the main sample.
- Outcome variable: Uptake in policy (1 if a scientific paper was cited by any policy document; 0 otherwise)
- Explanatory variables:
    - Disciplinary diversity of contributing authors (DDA)
    - Disciplinary diversity of cited references (DDR)
    - Control variables: normalised citation counts; normalised CiteScore; number of authors; number of countries, and average number of papers per author
- All models contain the same variables (although different transformations were applied to some variables in different models.
    - Model 1: logarithmic transformation of the explanatory variables.
    - Model 2: original form of the explanatory variables.
    - Model 3: normalised form of the explanatory variables (divided by one standard deviation). This model is the basis of the analysis regarding the size of link between cross-disciplinary research and policy uptake.

3.2.3. CHARACTERISATION OF OVERTON DATA SETS

Basic features of Overton records were also investigated and reported as descriptive statistics. Dimensions of interest here included what is commonly called a *citation peak*, or the average length of the period at which the greatest number of citations is received. If citation peaks are reached after long periods of time, then the utility of the approach in an evaluative context decreases. Another dimension is the shares of peer-reviewed publications cited at least once in the policy records, a figure that can validate or invalidate the relevance of the policy dimension in measuring societal outcomes more broadly.

Using a random sample of 50 FP-supported publications cited by documents indexed in Overton, Science-Metrix qualitatively assessed the extent to which such citations reflect research input into decision-making to address Question A. When some of these publications registered more than one citation from a policy document in Overton, only the first citation was assessed. The original, citing, policy document was retrieved and reviewed to (1) validate the existence of and locate the citation to the peer-reviewed publication of interest; (2) assess the overall character and content of the citing policy document (executive or legislative document; grey literature; affiliations of its authors; publishing organisation, etc.); and (3) assess referencing practices (format and presentation of citations made, as well as apparent motivations for making a citation) in the policy documents of interest. Finally, the quantitative data on the citation of FP7- and

H2020-supported papers in Overton's policy documents was used to describe the lag to policy citations and share of publications with at least one such citation.

4 Results & discussion

The analyses of the data set compiled for this study were divided into three main steps. First, the statistical model was estimated for a subset of roughly 34 000 FP-funded papers (i.e. supported through either FP7 or H2020 funding) with at least one UK-based author. By limiting the data set to UK papers, the goal was to avoid a possible measurement bias originating from differences in country coverage in Overton; Overton has a positive bias towards the UK. The results from this step, reported and analysed in Section 4.2.1, showed that this novel data source could be used in testing a fundamental assumption of programmes supporting cross-disciplinary research – namely, that this funding mechanism will fuel greater societal outcomes such as informing decision-making.

These initial results motivated the subsequent expansion of the analyses to a larger pool of FP-funded papers from any European country. This to assess the feasibility of using Overton in testing similar assumptions, but in a broader geographic context where differences in country coverage are known to exist. This also helped increase the relevance of the study's findings for the EC. We nevertheless acknowledge that not all European countries are equally represented in this broader data set. Indeed, the larger players (e.g. France and Germany, to name just a few) weigh heavily here and are still better covered in Overton than, say, the smaller Eastern European countries. In this second stage of analyses, the model was estimated once using the full set of about 126 000 FP-funded papers, and once using the 92 000 remaining FP-funded publications (i.e. those that did not include any author from the UK). Findings for the three data sets reported in Section 4.2.2 are broadly comparable, suggesting that the presented approach can be used in a broader European context. However, differences in the share of uptake across regions could be due to coverage issues in Overton confirming that it is best used in an investigative (i.e. answering broad research questions) than an evaluative (i.e. benchmarking) context. That said, it can still be used in benchmarking various programmes if their geographic/thematic scope is similar.

The third step consisted in presenting and describing some aggregated level data for the two FPs (i.e. FP7 and H2020), for the FP7 Specific Programmes (i.e. Euratom, Cooperation, Ideas, People and Capacities), and for selected FP7 calls for proposals (Section 4.3). These exploratory examples were provided to exemplify the subsequent use, as introduced in Section 3.1, that can be made of this study's indicators once evidence of a causal link between cross-disciplinary research and policy uptake has been established. The main rationale was that, since the indicators of cross-disciplinary research were revealed as suitable predictors for policy uptake in the statistical modelling of Section 4.2.1 and 4.2.2, they could be used in providing additional clues for the relative performance, in terms of policy uptake, of programmes promoting cross-disciplinary research to various degrees. This, as detailed in Section 3.1, would entail further qualitative analysis in a formal evaluation context to assist the interpretation of benchmarking against carefully chosen comparators.

Before diving into the results of the statistical models, some general statistics on the level of, and lag to, policy uptake are introduced in Section 4.1.

4.1 General statistics on the level of, and lag to, policy uptake

The qualitative assessment of policy document citations recorded by Overton found that 1 out of 50 policy document citations was a 'technical false positive' – that is, a mention to the cited publication could not be retrieved in the original citing policy document. The remaining 49 policy document citations were successfully validated in the original document. Of these, another 6 citations could be considered 'conceptual false positives,' in that the original citing documents were found to be scientific publications rather than grey literature documents or policy white papers, albeit copies of articles made available online on institutional websites rather than on journal websites. An additional 4 citations were made from 'hybrid' documents authored exclusively by academic authors but published by governmental or think tank organisations and whose content was judged to be very similar to that of a formal peer-reviewed research output. The remaining 39 citations originated from policy documents that must be considered part of the 'regulatory science' or science advisory branches of governance systems. Most of these documents appeared to have been authored by government scientists (sometimes in collaboration with academic scientists or scholars) and most consisted in syntheses and reviews of research findings. So while these citations should not be interpreted as indicative of advanced policy outcomes of research directly reaching the legislative or executive processes, they can be seen as achievements in contributing to the first stages of these processes, at the intersection between governance and academia. Within the 43 citations that can be considered valid, only 3 did not make a clear reference to the cited publication in the body of the document's text and make use of the reference to provide prior findings or support theory or method building; in only 6 cases was the reference made as part of a grouped citation containing multiple references (4 or more).

Prior work reported between 0.32 % and 1.41 % of publication sets being cited by at least one policy document (Bornmann et al., 2016; Haunschild & Bornmann, 2017; Tattersall & Carroll, 2018). Our results strongly contrast this work, with a figure of 6.0 % for the entire data set of FP-funded publications, 8.6 % for the subset limited to UK publications, and 5.1 % for the subset of non-UK publications.

Our own findings show much higher shares of publications with at least one policy citation than the shares reported in prior studies. These differences are most likely to be explained by differences in coverage between the Altmetric.com database, which archives altmetrics records from a broad range of platforms and may be more superficial in its coverage, and the more comprehensive Overton database, which focuses solely on policy documents. A more systematic comparison between coverage results is outside the scope of this brief; in addition, the data sets used in these studies have not been made openly available online.

The difference between our three analytical subgroups (UK, FP7-H2020 overall and non-UK) could reflect the coverage bias of Overton in favour of the United Kingdom.

While other studies could not provide robust data on intervals to policy citation peak (in terms of share cited), results from our global data set of FP7- and H2020-supported papers shows this peak may take place around the third year after publication year for 2008–2011 papers (dotted line, Figure 1); for 2012–2015 papers, the peak was between the second and third years after publication year (data not shown). Note that three years after publication year, about 50 % of the papers cited in policy had received their first citation (solid line, Figure 1). These findings hold true for the subset limited to UK papers. This suggests that policy altmetrics could find practical use in mid-term or ex-post (near the end as opposed to well after) programme evaluations, to a greater extent than previously envisaged.

KEY FINDING

Policy citation peaks were around the third year (for 2008–2011 papers) or between the second and third years (for 2012–2015 papers) after article publication year.

Such observations are positive signals that the quantitative analysis of policy citations could be relevant in programme evaluation exercises.

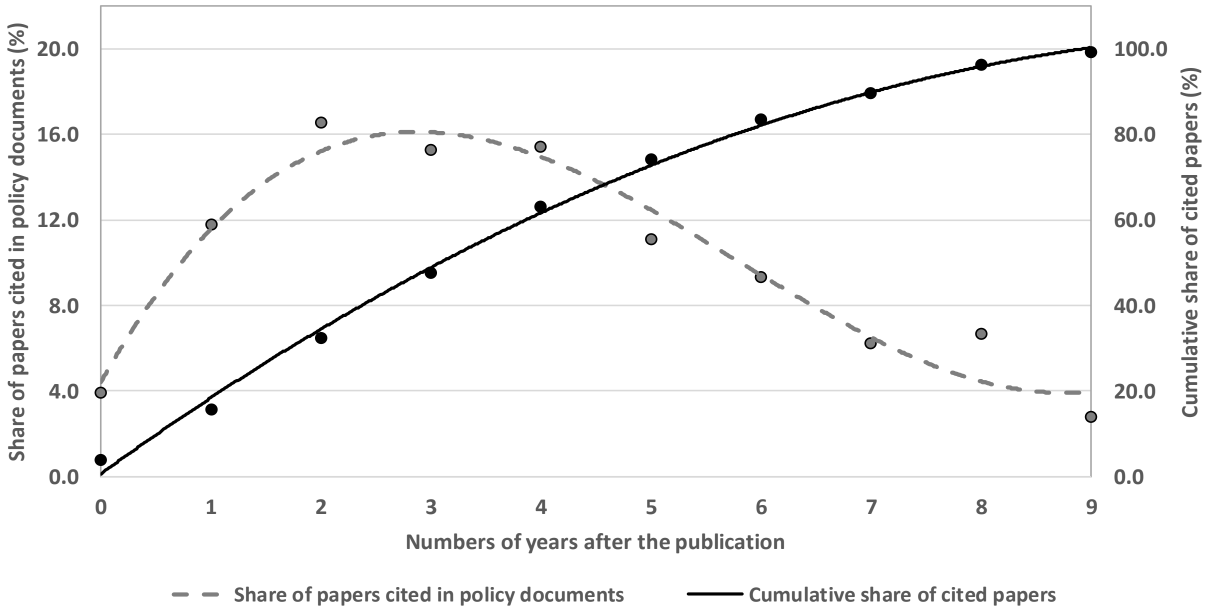

Figure 1 **Policy uptake of FP7- and H2020-funded papers by number of year(s) after publication, 2008–2011**

Source: Produced by Science-Metrix using Scopus (Elsevier) and Overton data

4.2 Modelling the link between cross-disciplinary research and policy uptake

4.2.1. FP-SUPPORTED PAPERS WITH AT LEAST ONE UK-BASED AUTHOR

Table 1 summarises the main results from statistical modelling for the sample of FP7- and H2020-funded papers with at least one UK-based author. Coefficients were mostly significant at an alpha of 0.01, while some were only so at 0.05. To simplify the assessment of effect sizes, we focus on Model 3. The last column in this table provides the effect on the probability of being cited following a change of one standard deviation on the explanatory variable if the probability of being cited (before the change on the explanatory variable) was at 10 %, which roughly corresponds to our sample's share of papers cited in policy documents.

Using this standard, we provide a brief summary for the DDA and DDR coefficients as they are the ones directly related to our core research question, the others being included as controls.

Results for DDA suggested that bringing together authors from different subfields of science is positively associated with policy uptake, with an associated increase in the probability of 1.2 percentage points for each additional standard deviation in this variable. DDR was not statistically associated with policy uptake after controlling for DDA, but it was without controlling for DDA (data not shown). The models, with or without publication year as a control, provided similar results (data not shown).

As others have found, these results highlight the multifaceted character of cross-disciplinary research and the difficulty in capturing all these dimensions with one or a few indicators (Q. Wang & Schneider, 2019).

KEY FINDING

Findings from the regression modelling of the association between policy citation and cross-disciplinary research (in addition to other dependent variables) show that DDA, but not DDR, appeared to foster policy citations. On average, the probability of being cited in policy documents increases roughly by 1.2 percentage points after a typical increase (1 standard deviation) in DDA. This is a potentially relevant increase, considering the average uptake of 6 % for the entire sample.

These results suggest that the collaborative dimension of cross-disciplinary research (captured through DDA) leads to results of a higher relevance for policymaking than the intellectual integration of findings from diverse origins alone (captured through DDR). Indeed, a single-author paper can cite diverse knowledge, yet this is intuitively less likely than with cross-disciplinary authorships. It could thus be hypothesised that cross-disciplinary research stemming from collaborative work (high DDA; the form typically promoted by funding programmes), which is not incompatible with high DDR, ensures a better integration of various stakeholders' considerations leading to socially relevant findings. Non-collaborative cross-disciplinary research (high DDR with low DDA) may instead lead to an increased intellectual complexity less favourable to policy uptake (see also Allmendinger, 2015). The former part of this hypothesis also raises the relevance of testing for cross-sectoral partnerships (papers involving governmental authors may have higher odds of policy uptake) in future work.

Table 1 Results of logistic regression for citation, in policy documents, of FP-funded publications with at least one UK-based author, 2008–2016

Variable	Logistic regression for variables in logarithmic form and in level: Log-transformed (Model 1)		Level (Model 2)		Level - for one standard dev. (Model 3)		
	Coefficients	Odds-ratio	Coefficients	Odds-ratio	Coefficients	Odds-ratio	Δp (baseline=10%)
Normalised citation score	0.960***	2.613	0.086***	1.09	0.613***	1.846	7.0%
	(0.035)		(0.006)		(0.043)		
Normalised CiteScore	0.024	1.024	0.153***	1.166	0.207***	1.23	2.0%
	(0.063)		(0.021)		(0.029)		
Document type (Review = 1)	0.727***	2.07	0.507***	1.661	0.507***	1.661	5.6%
	(0.101)		(0.096)		(0.096)		
Multidisciplinary (DDA)	0.234***	1.264	0.109***	1.115	0.129***	1.138	1.2%
	(0.058)		(0.026)		(0.031)		
Interdisciplinarity (DDR)	0.017	1.017	0.07	1.073	0.032	1.033	0.3%
	(0.125)		(0.089)		(0.041)		
Number of authors	-0.111**	0.895	-0.006***	0.994	-0.780***	0.458	-5.2%
	(0.048)		(0.001)		(0.176)		
Number of countries	0.163***	1.178	0.074***	1.077	0.279***	1.322	2.8%
	(0.063)		(0.011)		(0.040)		
Proportion of female authors	0.300**	1.35	0.152	1.164	0.033	1.033	0.3%
	(0.152)		(0.143)		(0.031)		
Avg. number of papers per author	0.107***	1.113	0.003***	1.003	0.115***	1.122	1.1%
	(0.040)		(0.001)		(0.039)		
Observations	37,897		37,897		37,897		
Log Likelihood	-3,446.58		-3,754.04		-3,754.04		
Wald Test (df = 11)	1,046.660***		528.910***		528.910***		
LR Test (df = 11)	1,345.617***		730.702***		730.702***		
Score (Logrank) Test (df = 11)	1,316.022***		766.147***		766.147***		

Note: Binary logarithms of normalised citation counts, normalised CiteScore, DDA, DDR, number of authors/countries, and average number of papers per author were used in Model 1. Therefore, the odds ratio of these coefficients refers to the variation in odds associated with a twofold change in the explanatory variables. *p<0.1; **p<0.05; ***p<0.01. Standard errors are reported in parenthesis.

Source: Produced by Science-Metrix using Scopus (Elsevier) and Overton data

4.2.2. BROADER SET OF FP-SUPPORTED PAPERS AS WELL AS THOSE WITHOUT ANY UK-BASED AUTHOR

Table 2 expands the previous analysis, reporting on the findings from statistical models that followed the same specification as Model 3 from Table 1, but for two additional sets of papers:

- Europe (Model 4), corresponding to all FP-supported papers, regardless of the authors' affiliation countries, and

- Europe non-UK (Model 5), corresponding to all papers in the data set except those having at least one UK-based author (i.e. those in the data set for Model 3).

The coefficients for DDA and DDR showed the same signs and statistical significance in the three models. However, the lower coefficient observed for DDA in Model 5 (non-UK authors) suggests that multidisciplinary collaboration had less importance in driving policy uptake for non-UK publication output, or that the coefficient in this model was affected by the lower coverage of Overton outside the UK. This latter hypothesis may very well be at play considering the lower share of papers cited in Overton for non-UK papers (5.1 %) compared to UK papers (8.6 %). Accordingly, the results suggest that relying on Overton data to rank countries for policy uptake could lead to unfair comparisons. On the other hand, the possible coverage biases do not appear to be of a sufficient magnitude to prohibit the use of Overton in investigative rather than in summative programme evaluations in regions outside the UK, as was done here. Furthermore, the results also suggest that tracking scientific papers in policy databases could still represent a viable strategy to measure the policy uptake of research findings across different projects or programmes once differences in the distribution of papers across countries has been accounted for.

The coefficients of most of the remaining variables were also comparable across the three data sets used in Table 2. In two cases (number of authors and average number of papers per author), the coefficients in the non-UK data set were no longer statistically different from 0. These two indicators were included as control variables with no prior expectations regarding their signs. As with DDA, these differences may have reflected cross-countries differences in the coefficients or may have resulted from differences in coverage among different countries. The fact that none of the coefficients presented different signs in these different data sets pointed to some degree of robustness in this indicator as one way to capture the policy uptake of publications.

KEY FINDING

In country-level comparisons where non-English-speaking countries are involved, uneven distribution of articles across countries should be considered when measuring levels of citations from policy documents (as recorded, for example, in the Overton database).

Table 2 Results of logistic regression for citation, in policy documents, of FP-funded publications overall and by subset, 2008–2016

Variable	Logistic regression: standardised variables - UK, Europe, Non-UK								
	UK			Europe			Europe non-UK		
	Coefficients	Odds-ratio	Δp (baseline=10%)	Coefficients	Odds-ratio	Δp (baseline=10%)	Coefficients	Odds-ratio	Δp (baseline=10%)
Normalised citation score	0.613*** (0.043)	1.846	7.0%	0.604*** (0.022)	1.829	6.9%	0.655*** (0.028)	1.925	7.6%
Normalised CiteScore	0.207*** (0.029)	1.230	2.0%	0.146*** (0.015)	1.157	1.4%	0.101*** (0.021)	1.106	0.9%
Document type (Review = 1)	0.507*** (0.096)	1.660	5.6%	0.722*** (0.054)	2.059	8.6%	0.813*** (0.071)	2.255	10.0%
Multidisciplinary (DDA)	0.129*** (0.031)	1.138	1.2%	0.101*** (0.016)	1.106	0.9%	0.065*** (0.021)	1.067	0.6%
Interdisciplinarity (DDR)	0.032 (0.041)	1.033	0.3%	0.021 (0.021)	1.021	0.2%	0.015 (0.025)	1.015	0.1%
Number of authors	-0.780*** (0.176)	0.458	-5.2%	-0.635*** (0.076)	0.530	-4.4%	-0.008 (0.018)	0.992	-0.1%
Number of countries	0.279*** (0.040)	1.322	2.8%	0.251*** (0.018)	1.285	2.5%	0.106*** (0.017)	1.112	1.0%
Proportion of female authors	0.033 (0.031)	1.033	0.3%	0.021 (0.016)	1.021	0.2%	0.022 (0.020)	1.022	0.2%
Avg. number of papers per author	0.115*** (0.039)	1.122	1.1%	0.052*** (0.020)	1.053	0.5%	0.035 (0.025)	1.036	0.3%
Observations	37,897			137,419			99,522		
Log Likelihood	-3,754.04			-12,927.00			-7,757.00		
Wald Test (df = 11)	528.910***			1,836.000***			1,005.000***		
LR Test (df = 11)	730.702***			2,474.000***			1,413.000***		
Score (Logrank) Test (df = 11)	766.147***			2,714.000***			1,411.000***		

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are reported in parenthesis.
Source: Produced by Science-Metrix using Scopus (Elsevier) and Overton data

4.3 Descriptive analyses of policy uptake and cross-disciplinarity by subfield and funding mechanism

Within the context of programme evaluation, the questions could be whether the evaluated group scores higher than comparators for its shares of publications cited by policy documents, and to what extent the observed differences can be attributed to the differences in multidisciplinarity between them (which could themselves result from differences in the degree to which their corresponding funding mechanisms fostered multidisciplinarity). The statistical modelling analysis, as reported in the previous section, will not directly address this question. Those models, by accounting for fixed characteristics of the various research projects in the data set, use only within-project variation of the included indicators and, therefore, are more suited to exploring an average impact of cross-disciplinarity on policy uptake. Nevertheless, their results should motivate the use of the proposed indicators in evaluation of individual programmes, by comparing their scores with those of carefully chosen comparators as introduced in our proposed methodological framework (Section 3.1).

In brief, once an assumption underlying many funding programmes has been verified, as in Section 4.2, it becomes easier to assess the contribution of a specific programme building on that assumption to deliver its desired outcome – here cross-disciplinarity influencing policy uptake – even when the volume of data linked to the programme does not enable the highly controlled modelling presented above.

Programme evaluations could then consider performance on the indicators of interest (i.e. cross-disciplinarity, policy uptake and other controls) – along with the typical qualitative analysis presented in case studies (e.g. to rate the intensity with which cross-disciplinarity is enforced) – to assess whether funding schemes promoting multidisciplinarity may have led to higher levels of policy uptake in specific research programmes. In theory, there should be a positive correlation between all these dimensions, everything else being held constant (which we admit is difficult to achieve). In these case studies, the choices of comparators should thus, in addition to some variables included in the previous models, account for the qualitative characteristics that are intrinsic to each research programme, such as characteristics of researchers, that would have exerted impact on policymaking even in the absence of the funding.

The following tables present the scores for variables used in the statistical models. In addition to share of papers cited in policy documents, a normalised version of this share is also presented. Normalisation accounted for differences in the patterns of policy uptake across scientific subfields and years (i.e. the variable is divided by the mean across each subfield and year). The data set of FP-funded papers was used as a baseline for this procedure since, for this study, the policy data were not available beyond that sample. This normalisation was therefore implemented differently than for indicators such as DDA, DDR and average of relative citations (ARC)[2], for which the entire Scopus data set was available. Still, it enabled adequate comparisons at the programme (FP7 vs H2020), specific programme (for FP7 only) and call level.

Table 3 first presents the most voluminous subfields in the FP-funded (both FP7 and H2020) publication set, and those with the highest shares of papers cited in policy documents. The results show that many areas in the social sciences and humanities (e.g. economics, political science and public health) achieve much higher rates of policy uptake than some 'hard sciences' subfields such as general physics. These findings converge with prior results on the distribution of altmetrics citations across disciplines and subfields.(Haustein, Costas, & Larivière, 2015) As a result, research programmes focusing on the former relative to the latter should be expected to, on average, achieve higher rates of policy uptake, illustrating the need to normalise policy uptake by subfield

[2] The ARC is built as an average of paper citation scores normalised by subfield, year and document type.

before comparing different programmes. These subfield-level results subsequently served to compute the normalised share of papers cited in policy documents, minimising disciplinary biases across different research programmes.

Table 3 Shares of FP-funded publications cited in policy documents by subfield, 2008–2016

Top 10 most frequent subfields in the data set			Top 10 subfields most cited in policy in the data set		
Subfield	Share cited in policy	Number of Publications	Subfield	Share cited in policy	Number of Publications
Developmental Biology	3.2%	6,581	Economics	72.0%	529
Applied Physics	0.4%	6,360	Political Science & Public Administratior	41.9%	136
Nuclear & Particle Physics	0.3%	5,338	Toxicology	38.8%	642
Astronomy & Astrophysics	0.6%	4,944	Science Studies	35.9%	132
Neurology & Neurosurgery	2.2%	4,444	Health Policy & Services	34.8%	217
General Physics	0.5%	4,022	Public Health	33.7%	404
Nanoscience & Nanotechnology	1.4%	3,977	Geography	33.5%	165
Meteorology & Atmospheric Sciences	30.0%	3,837	General & Internal Medicine	33.0%	515
Chemical Physics	0.8%	3,649	Environmental Sciences	31.4%	1,025
Fluids & Plasmas	1.6%	3,451	Economic Theory	31.2%	101

Source: Produced by Science-Metrix using Scopus (Elsevier) and Overton data

Table 4 summarises observations computed separately for each Framework Programme (i.e. FP7 and H2020). The stand-out difference between the publications of each group from the two FPs was in the variable share cited in policy (6.2 % for FP7 vs 3.6 % for H2020), a difference that vanishes completely in the normalised version of the indicator (1.00 for FP7 vs 1.01 for H2020). Differences in the raw share cited in policy could be explained by differences in programme periods. The more recent H2020-funded publications, relative to FP7-funded papers, likely had lower odds of having been cited by policy documents at the time this study was completed. According to the normalised version of the indicator by subfield, but also by year, both programmes thus appear to have performed similarly in terms of influencing decision-making.

The publications from FP7 accounted for more than 90 % of the data set. This is not surprising as this programme closed in 2013 and was followed by H2020, for which less than half of the funding period (2014–2020) was considered here. This is because the analysis period was restricted until 2016 to enable enough time for the uptake of publications in policy documents[3]. The remaining analyses thus focus on FP7 only.

Table 4 Shares of publications cited in policy documents, along other relevant indicators, by Framework Programme, 2008–2016

Framework programme	N	Share cited in policy	Norm share cited in policy	DDA	DDR	ARC	Proportion of reviews
FP7	117,016	6.2%	1.00	1.17	1.02	2.32	6.2%
H2020	10,189	3.6%	1.01	1.12	0.99	2.40	6.5%

Source: Produced by Science-Metrix using Scopus (Elsevier) and Overton data

Table 5 summarises observations for the five FP7 Specific Programmes. Note that the two highest DDA and DDR scores (for Euratom and Cooperation) are from the two programmes that also displayed the highest normalised shares of publications cited in

[3] This finding also highlights how the previous modelling exercises relied mostly on FP7 projects. A similar exercise could be performed in the future on a larger data set of H2020 publications for additional validation (a preliminary test based strictly on the current H2020 subsample suggested comparable results to those reported in Section 4.2 for the coefficient of DDA).

policy documents. This is certainly not enough evidence to claim any sort of causality. However, the above modelling already demonstrated a correlation that accounted for many types of potential confounders and could be seen as an estimate that approximates the real causal relationship (at least compared to measures of simple correlations). Therefore, the higher level of cross-disciplinarity observed for these two specific programmes could have played a role in this higher uptake.

As already shown in the modelling of Section 4.2, other factors are also at play, such as scientific impact (ARC; normalised citation in Section 4.2). Here, the highest scientific impact score was observed for Ideas, which had the second-to-last score for normalised policy uptake and one of the lowest DDA and DDR scores. On the other hand, Cooperation, which was second for impact and DDA/DDR, was also second for policy uptake. The many factors at play could also explain the considerably lower than average score in normalised policy uptake for Capacities, which also scored above the FP7 average for DDA. For instance, since Capacities promoted the development of shared research infrastructures and enhanced the innovative capacities of small and medium enterprises, its publications – while requiring input from a diverse knowledge base – might not be tackling issues (such as the SDGs) of the highest relevance to policymaking compared to Cooperation and Euratom, which show strong connections to the SDGs. This further emphasises the need to incorporate qualitative evidence in interpreting observed differences between comparators in the context of formal programme evaluations building on pre-existing evidence supporting programmatic assumptions. Additionally, variability across calls and projects within individual programmes could help further explain observed differences at higher aggregation levels.

Table 5 Shares of publications cited in policy documents, along other relevant indicators, by FP7 Specific Programme, 2008–2016

Specific Programmes	N	Share cited in policy	Norm share cited in policy	DDA	DDR	ARC	Proportion of reviews
FP7	**117,016**	**6.2%**	**1.00**	**1.17**	**1.02**	**2.32**	**6.2%**
Capacities	8,437	3.6%	0.47	1.19	1.02	1.72	3.6%
Cooperation	56,605	9.3%	1.28	1.27	1.05	2.32	7.2%
Euratom	576	10.7%	9.06	1.33	1.09	1.45	4.2%
Ideas	40,536	3.8%	0.69	1.09	0.99	2.60	5.8%
People	19,023	3.1%	0.73	1.05	0.97	2.02	4.8%

Source: Produced by Science-Metrix using Scopus (Elsevier) and Overton data

Table 6 presents the data aggregated by selected FP7 calls for proposals. Calls with the 10 highest and the 10 lowest shares of papers cited in policy (before the normalisation) are presented. Among the calls with the highest policy uptake, nine scored higher than the average DDA for FP7 (i.e. > 1.17), while only four did so for those calls with the least policy uptake. Similarly, seven of the calls with the highest policy uptake scored above the FP7 average in scientific impact (i.e. ARC > 2.32), while only four did so for those calls with the least policy uptake. At this aggregation level, the relationships observed at paper level in the statistical models (Section 4.2) appear more clearly, highlighting the relevance of accounting for within-programme variation in programme evaluations.

KEY FINDING

At the level of FP7 and H2020 programmes, policy citation levels generally (but not always) correlated to either recorded DDA levels or prior expectations given each programme's goals. These findings indicate that DDA is not the sole research practice that fostered policy uptake of research findings, and that multiple research and engagement practices are combined in processes of evidence-based policymaking. In the absence of a fuller picture, however, policymakers and funding agency managers should almost assuredly add in provisions to support collaborative multidisciplinarity when designing funding programmes, if the fostering of policy-related outcomes of research are deemed an important component of these programmes.

The discordant cases (i.e. calls highly cited in policy documents with DDA and/or ARC below the FP7 average and low cited calls with DDA and/or ARC above the FP7 average) suggests that other factors may also be important in explaining observed differences in performance across different calls. For example, the FP7-OCEAN-2010 call (The Ocean of Tomorrow) scored highly in terms of policy uptake, while scoring a DDA and ARC below the FP7 average. However, it is highly relevant to the study of climate change under FP7 Cooperation, which is a topic of great interest within the policy sphere. On the other hand, the FP7-NMP-2008-LARGE-2 call (Nanosciences, Nanotechnologies, Materials and new Production) scored low in terms of policy uptake, while scoring a DDA and ARC well above the FP7 average. It follows quite naturally that this call performed well in DDA and ARC since the nanosciences and nanotechnologies help bring to life many novel solutions to applied problems in diverse fields. However, the specific solutions/technologies they enable are much less likely to capture the interest of a broad range of policymakers than research on climate change. In this latter case, uptake in patents would be more relevant. In fact, the connection between interdisciplinarity and innovation has also been previously studied to test the assumption that the former fuels the latter, an assumption that has also been used in designing various funding programmes (Campbell et al., 2017). These findings reinforce the previous remarks on the relevance of building a body of qualitative evidence to complement the insights from these indicators in formal programme evaluations, as well as the importance of carefully selecting appropriate comparators (e.g. similar topics, similar funding mechanisms).

Table 6 Shares of publications cited in policy documents, along other relevant indicators, by FP7 calls for proposals, 2008–2016

Call for proposals	N	Share cited in policy	Norm share cited in policy	DDA	DDR	ARC	Proportion of reviews
10 highest policy rate							
FP7-SSH-2007-1	201	58.2%	3.41	1.29	0.97	3.04	2.5%
FP7-ENV-2011	486	39.8%	1.73	1.29	1.03	3.03	5.0%
FP7-OCEAN-2010	133	37.6%	1.72	1.12	1.05	1.69	4.5%
FP7-ENV-2007-1	1,636	33.7%	2.08	1.35	1.09	2.52	5.1%
FP7-ENV-2010	977	32.6%	1.39	1.27	1.04	2.64	5.2%
FP7-ENV-2012-two-stage	296	31.5%	2.28	1.21	0.96	2.21	5.9%
FP7-ENV-2009-1	1,083	31.3%	1.76	1.31	1.03	2.32	5.4%
FP7-ENV-2008-1	1,815	31.1%	1.47	1.23	0.98	2.34	4.5%
ERC-2010-AdG_20100407	318	30.2%	1.80	1.58	1.06	3.33	4.7%
FP7-ENV-2013-two-stage	216	30.1%	1.35	1.21	0.97	2.92	5.5%
10 lowest policy rate							
FP7-ICT-2009-6	876	0.9%	0.50	1.26	1.03	1.97	5.3%
FP7-INFRASTRUCTURES-2011-1	691	0.9%	0.14	0.96	0.86	1.73	4.0%
FP7-ENERGY-NMP-2008-1	138	0.7%	0.28	1.35	1.18	3.02	2.2%
FP7-ICT-2013-C	138	0.7%	0.45	0.90	0.97	2.24	0.7%
FP7-ICT-2011-C	283	0.7%	0.80	1.04	0.98	1.98	3.9%
FP7-PEOPLE-IIF-2008	149	0.7%	0.02	1.09	0.96	1.58	1.3%
ERC-2012-StG_20111012	1,528	0.6%	0.49	0.95	0.94	2.88	2.6%
FP7-ICT-2013-FET-F	435	0.5%	0.26	1.10	1.01	2.78	3.9%
FP7-PEOPLE-2007-4-3-IRG	326	0.3%	0.06	1.36	1.11	1.58	4.6%
FP7-NMP-2008-LARGE-2	558	0.0%	0.00	1.50	1.09	2.39	5.4%

Source: Produced by Science-Metrix using Scopus (Elsevier) and Overton data

Table 7 presents results of regressions performed at the level of research projects. They include fixed effects for calls for proposals, therefore accounting for unobserved characteristics of each call. However, relevant unobserved characteristics of research projects, such as the composition of each research team, could not be accounted for. Most of the coefficients for the second model using the normalised policy uptake (Table 7) were consistent with those computed at paper level in Section 4.2. For instance, DDA again appeared as a relevant variable to explain variations in the share of papers cited in policy documents.

These findings were interesting in revealing that an association between DDA and policy uptake could also be observed at the level of research projects. Future work could substantiate these findings by using more advanced techniques such as hierarchical modelling and by accounting for the presence of a large proportion of zeros among the variable for the share of publications cited in policy documents. In any case, these results based on indicators measured at project level confirm those obtained at paper level from the previous section. It thus appears highly relevant to support the use of these indicators in situations where they are measured at higher levels of aggregation, such as funding programmes, provided comparators are adequately chosen and qualitative inputs are used in interpreting observed differences.

Table 7 Relationship between cross-disciplinary research, plus other controls, and uptake in policy documents measured at project level, 2008–2016

Variable	Raw rate of policy citation	Normalised rate of policy citation
Normalised citation score	0.005***	0.103***
	(0.002)	(0.036)
Normalised CiteScore	0.007	0.475***
	(0.007)	(0.174)
Document type (Review = 1)	-0.036	-0.93
	(0.042)	(1.002)
Multidisciplinary (DDA)	0.023***	0.673***
	(0.007)	(0.160)
Interdisciplinarity (DDR)	-0.008	-0.462
	(0.013)	(0.312)
Number of authors	-0.0003***	-0.004**
	(0.0001)	(0.002)
Number of countries	0.017***	0.209***
	(0.003)	(0.063)
Proportion of female authors	0.091***	0.234
	(0.024)	(0.582)
Avg. number of papers per author	-0.0002**	-0.002
	(0.000)	(0.003)
Observations	1,708	1,706
R^2	0.071	0.045
F Statistic	12.762*** (df = 9; 1513)	7.917*** (df = 9; 1511)

Note: The models were estimated using linear regression at project level with fixed effects for calls for proposals.
*p<0.1; **p<0.05; ***p<0.01
Source: Produced by Science-Metrix using Scopus (Elsevier) and Overton data

5 Conclusions and limitations

5.1 Discussion of main findings (Questions A, B and C)

While prior studies using quantitative analysis of policy citations towards research publications saluted the comparatively sound conceptual basis of using this type of analysis to monitor some of the societal outcomes of research, they found cause for caution in the current infrastructure available for its implementation. Here, the feasibility of this altmetrics approach to the quantitative measurement of societal outcomes of research was assessed using a large data set of publications (~137 000) resulting from FP7 and H2020 projects, and Overton – a novel database intending to exclusively record the policy-related outcomes of research.

Following a preliminary quantitative and qualitative assessment of Overton data in this context (Question A), the new database appears to be an important addition to the quantitative toolbox of altmetrics and other instruments for tracking such societal research outcomes. At least one policy citation could be found for as many as 6 % of publications from these EU-funded research projects using Overton, a much higher figure than reported in any of the previous studies using alternative data sources, although the publication sets used were admittedly constructed quite differently in each study (Bornmann et al., 2016; Haunschild & Bornmann, 2017; Tattersall & Carroll, 2018). In fact, this figure was as high as 42 % and 72 % in some subfields (Political Science & Public Administration; and Economic). This conclusion is also supported by the observation of numerous citation peaks between years 2 and 3 after publication, which means this data source shows potential for informing decision-making in a timely manner. Additionally, in an ancillary qualitative evaluation of the reliability of Overton data, it was found that the number of false positives in that database should be low and that motivations behind citation acts were generally clear and convincing.

The assumption that cross-disciplinary research is more likely to foster broad societal outcomes than disciplinary research has been highly prevalent in current policymaking and research, but there is surprisingly little work that directly tests this relationship. Regression analysis using the above sample of FP7 and H2020 publications showed that higher policy uptake was correlated with collaborative cross-disciplinary research (DDA) but not with the intellectual integration of disciplines, narrowly defined (DDR), once DDA had been accounted for (Question B). We contend that our positive findings for Question A and Question B provide a strong basis for selecting instruments that implement DDA when designing research funding programmes with an explicit goal to increase societal outcomes, in the form of increased knowledge transfer towards policymakers (Question C). We also find that scientific citations correlate quite strongly with citations from policy-related documents, indicating that policymakers may rely on traditional markers of excellence when seeking out scientific evidence to support their activities, or that work with a higher relevance for policy may also tend to be more highly cited. Reverse causality is also a relevant hypothesis for the strong correlation observed between scientific and policy citations. In many cases, citation in policy could even have preceded scientific citations, included in the models deployed here mainly to control for a possible indirect link between cross-disciplinarity and citation in policy through a higher impact within the scientific community.

5.2 Recommendations for the use of policy document citations in future studies and programme evaluations (Question D)

When investigating the policy outcomes of a funding instrument (e.g. a programme specifically supporting DDA), researchers often face small numbers of projects and publications, which may mean that it is not possible to unambiguously attribute observed changes to the given instrument; in other words, they may have insufficient data to robustly estimate a statistical model controlling for confounders (i.e. with counterfactual). The paper-level approach used here to test the association between cross-disciplinarity and policy citations was implemented on a much larger set of publications than any of these specific programmes would provide. With the association between DDA and policy uptake validated at the paper level on a large scale, we argue that papers funded through specific programmes funding cross-disciplinary teams would be more likely, in the aggregate, to impact policy to a degree proportional to the effectiveness of the incentives to integrate members from different disciplines in research projects.

Accordingly, the reliability of conclusions from evaluations using relatively simple analyses[4] and reporting positive societal outcomes (in the form of policy uptake) from programmes specifically promoting DDA could be deemed as more reliable if they were building on prior validation – on a large scale not specific to any such programmes – of the assumption that the promoted mechanism (e.g. DDA) can indeed act as a driver of policy outcomes (as provided here under the findings for Questions A, B and C, above). For example, a specific programme could be evaluated by simple bibliometric benchmarking of DDA and policy citations against other programmes with similar (possibly to a different extent) mechanisms for fostering DDA. In such a context, the

[4] For example, analyses suffering some limitations such as benchmarking of DDA and policy uptake against other funders without control group, a before/after analysis without (or with an imperfect) a control group, or lack of sufficient statistical power to derive robust conclusions.

evaluation should also rate the intensity[5] with which that programme aims to foster DDA relative to the chosen comparators (for example, some programmes may have mandatory requirements for diverse teams, while others may encourage such teams without enforcing them). With evidence of a general and positive link between DDA and policy uptake, programmes with mechanisms more strongly geared towards DDA should, at the aggregate programme level, perform better on indicators of DDA and policy uptake. Accordingly, if an increased level of DDA and policy uptake is observed for a programme's supported scholars (even without a control group), and if the levels of DDA achieved appear (through qualitative appraisal) to be fostered by the programme's design and associated instruments, evaluators could more safely conclude a positive outcome of the programme.

5.3 A general framework for investigating the societal outcomes of policy interventions (Question E)

Few examples are currently available of formal evaluations using altmetrics or quantitative societal outcomes data (with the exception of patent-related metrics), demonstrating a general caution towards the use of altmetrics in programme evaluation contexts. We propose that the paper-level approach employed here can be extended to the evaluation of other societal outcomes of research that are realised at least in part through peer-reviewed publications. In situations where the attribution of changes in practices and outcomes to a given policy instrument would be hampered by small volumes of observations or technical shortcomings such as those commonly observed in altmetrics methods, large-scale testing of the mechanism that links together specific paper features with a set of societal outcomes can serve as a foundation for more pointed follow-up investigations, whether in research or evaluative contexts.

5.4 Limitations

Going forward, it will be possible to deploy the Overton database as well as the research strategies implemented here in two core contexts: in quantitative research on cross-disciplinary research, societal outcomes of research and/or the development of altmetrics, and in the applied context of (research and innovation) programme evaluation. In the latter context, we advocate for policy citation analysis to be used in an investigative manner, to answer well-defined research questions that elucidate major mechanisms of action for the programme under review. More work is needed to characterise these tools before they can provide evidence for proper benchmarking exercises. For instance, our results are subject to a number of limitations:

- Citations towards journal articles from policy documents capture only a subset of science–policy interactions, with prior reports showing that most knowledge transfer takes place through tacit and local engagement rather than formal channels; nevertheless, the high shares of policy citedness among EU-funded papers reported here may indicate that the importance of formal channels for knowledge transfer towards policy has been underestimated.

- The Overton database has yet to be the subject of sustained investigation in the altmetrics and bibliometrics communities; further work is necessary to better understand the limitations of this data set. Particularly, the citations from policy documents retrieved for the set of publications examined originated to a large extent from regulatory science or scientific advisory documents rather than executive or legislative documents. This observation indicated that although some societal impact had been achieved by the peer-reviewed publications examined, this impact was located very much in the first steps of the evidence-based policymaking process, rather than in the deeper stages of integration. Do note that Overton does contain

[5] From a qualitative assessment relying on, for example, a document review, interviews and/or surveys of relevant stakeholders.

records on executive and legislative documents, and that our finding may be a function of the specific publication set used here.

- The binary indicator used to represent citation in policy documents does not capture the differences in the number of citations received by scientific papers from policy documents. Papers are treated similarly whether they have been cited only once or many times in policy documents. This option was shaped by the low proportion of papers being cited in policy and by the observation that the number of citations received may be a less precise indicator compared to the binary variable chosen, especially in a database that has not been frequently used before in such work.

- Causal claims about regression coefficients based on observational data sets are usually unlikely. Nevertheless, the models reported here accounted for different confounders and for fixed effects for research projects. It should help to approximate these coefficients to the actual causal relationship, compared to simple measures of correlations that do not account for the effect of confounders. However, the closeness of the reported coefficients and the true causal relationships is hard to assess. Triangulation with future work, quantitative or qualitative, should help to validate the findings reported in this paper.

- Policymakers and funding agency managers may be interested in knowing if there is an optimal configuration of research teams that promotes the production of findings with high likelihood of policy relevance. The DDA indicator is perhaps more sensitive to the distance between disciplines than the number of disciplines, however. Further work with this indicator will be necessary to better understand its behaviour and how it reacts to the diversity of research teams encountered in practice, in the field. Any work that examines research team configuration is also likely to require in-depth, qualitative as well as quantitative evidence on practices in the field and be very sensitive to the role of varying institutional contexts.

REFERENCES

Agresti, A. (2012). 16.1 Delta Method. In *Categorical Data Analysis* (3rd ed.). New York: John Wiley & Sons.

Allmendinger, J. (2015). *Quests for interdisciplinarity: A challenge for the ERA and HORIZON 2020. Policy Brief by the Research, Innovation, and Science Policy Experts (RISE)*. Retrieved from https://ec.europa.eu/research/openvision/pdf/rise/allmendinger-interdisciplinarity.pdf

Bornmann, L., Haunschild, R., & Marx, W. (2016). Policy documents as sources for measuring societal impact: how often is climate change research mentioned in policy-related documents? *Scientometrics*, *109*, 1477–1495. https://doi.org/10.1007/s11192-016-2115-y

Bozeman, B., Fay, D., & Slade, C. P. (2012). Research collaboration in universities and academic entrepreneurship: the-state-of-the-art. *The Journal of Technology Transfer*, *38*(1), 1–67. https://doi.org/10.1007/s10961-012-9281-8

Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, *534*(7609), 684–687. https://doi.org/10.1038/nature18315

Buenstorf, G., & Koenig, J. (2020). Interrelated funding streams in a multi-funder university system: Evidence from the German Exzellenzinitiative. *Research Policy*, *49*(3), 103924. https://doi.org/10.1016/j.respol.2020.103924

Campbell, D., Deschamps, P., Côté, G., Roberge, G., Lefebvre, C., & Archambault, É. (2015). Application of an “interdisciplinarity” metric at the paper level and its use in a comparative analysis of the most publishing ERA and non-ERA universities. *20th International Conference on Science and Technology Indicators*. Retrieved from http://science-metrix.com/sites/default/files/science-metrix/publications/campbell_et_al_sti2015_short_paper_final_web.pdf

Campbell, D., & Struck, B. (2019). Reliability of Scopus author identifiers (AUIDs) for research evaluation purposes at different scales. *17th International Conference of the International Society for Scientometrics and Informetrics (ISSI), 2–5 September 2019, Proceedings Vol. II*, 1276–1287. Retrieved from http://issi-society.org/publications/issi-conference-proceedings/proceedings-of-issi-2019/

Campbell, D., Struck, B., Tippett, C., & Roberge, G. (2017). Impact of multidisciplinary research on innovation. *16th International Conference of the International Society for Scientometrics and Informetrics (ISSI), 16–20 October 2017*. Retrieved from http://www.science-metrix.com/sites/default/files/science-metrix/publications/issi2017_paper_153_d_campbell_impact_multidisciplinarity.pdf

Chavarro, D., Tang, P., & Rafols, I. (2014). Interdisciplinarity and research on local issues: evidence from a developing country. *Research Evaluation*, *23*(3), 195–209. https://doi.org/10.1093/reseval/rvu012

Deste, P., Mahdi, S., Neely, A., & Rentocchini, F. (2012). Inventors and entrepreneurs in academia: What types of skills and experience matter? *Technovation*, *32*(5), 293–303. https://doi.org/10.1016/j.technovation.2011.12.005

Dworkin, J. D., Shinohara, R. T., & Bassett, D. S. (2019). The emergent integrated network structure of scientific research. *PLoS ONE*, *14*(4). https://doi.org/10.1371/journal.pone.0216146

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from national systems and “Mode 2” to a triple helix of university–industry–government relations. *Research Policy*, *29*(2), 109–123. Retrieved from http://www.sciencedirect.com/science/article/pii/S0048733399000554

European Commission Expert Group of Altmetrics. (2017). *Next-generation metrics: Responsible metrics and evaluation for open science*. Brussels.

European Research Council. (2018). *Qualitative evaluation of completed projects funded by the European Research Council*. Retrieved from https://erc.europa.eu/sites/default/files/publication/files/ERC_Science_behind_the_projects_FP7-2007-2013.pdf

Garner, J. G., Porter, A. L., Newman, N. C., & Crowl, T. A. (2012). Assessing research network and disciplinary engagement changes induced by an NSF program. *Research Evaluation*. https://doi.org/10.1093/reseval/rvs004

Gleed, A., & Marchant, D. (2016). *Interdisciplinarity survey report for the Global Research Council 2016 annual meeting*. Stockport.

Haunschild, R., & Bornmann, L. (2017). How many scientific papers are mentioned in policy-related documents? An empirical investigation using Web of Science and Altmetric data. *Scientometrics*, *110*(3), 1209–1216. https://doi.org/10.1007/s11192-016-2237-2

Haustein, S. (2016). Grand challenges in altmetrics: heterogeneity, data quality and dependencies. *Scientometrics*, *108*(1), 413–423. https://doi.org/10.1007/s11192-016-1910-9

Haustein, S., Bowman, T. D., & Costas, R. (2016). Interpreting “altmetrics”: viewing acts on social media through the lens of citation and social theories. In C. R. Sugimoto (Ed.), *Theories of Informetrics: A Festschrift in Honor of Blaise Cronin*. Retrieved from http://arxiv.org/abs/1502.05701

Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing Social Media Metrics of Scholarly Papers: The Effect of Document Properties and Collaboration Patterns. *PLOS ONE*, *10*(3), e0120495. https://doi.org/10.1371/journal.pone.0120495

Hird, M. D., & Pfotenhauer, S. M. (2017). How complex international partnerships shape domestic research clusters: Difference-in-difference network formation and research re-orientation in the MIT Portugal Program. *Research Policy*, *46*(3), 557–572. https://doi.org/10.1016/j.respol.2016.10.008

Koier, E., & Horlings, E. (2015). How accurately does output reflect the nature and design of transdisciplinary research programmes? *Research Evaluation*, *24*(1), 37–50. https://doi.org/10.1093/reseval/rvu027

Langfeldt, L., & Scordato, L. (2015). *Assessing the broader impacts of research. A review of methods and practices*. Retrieved from Nordic Institute for Studies in Innovation, Research and Education website: https://nifu.brage.unit.no/nifu-xmlui/handle/11250/282742

Larivière, V., Bowman, T., Robitaille, J.-P., Pollitt, A., Gunashekar, S., & Macaluso, B. (2015). *Comparative scientometric assessment of the results of ERC funded projects. Alternative metrics report (D7)*. Brussels.

LERU. (2016). *Interdisciplinarity and the 21st century research-intensive university*. Retrieved from https://www.leru.org/files/Interdisciplinarity-and-the-21st-Century-Research-Intensive-University-Full-paper.pdf

Lyall, C., Bruce, A., Marsden, W., & Meagher, L. (2013). The role of funding agencies in creating interdisciplinary knowledge. *Science and Public Policy*, *40*(1), 62–71. https://doi.org/10.1093/scipol/scs121

Lyall, C., Meagher, L., & Bruce, A. (2015). A rose by any other name? Transdisciplinarity in the context of UK research policy. *Futures*, *65*, 150–162. https://doi.org/10.1016/j.futures.2014.08.009

Molas-Gallart, J., Rafols, I., & Tang, P. (2014). On the relationship between interdisciplinarity and impact : different modalities of interdisciplinarity lead to different types of impact. *Journal of Science Policy and Research Management*, *29*(September), 69–89.

National Information Standards Organization. (2016). *Outputs of the NISO Alternative Assessment Metrics Project*. Retrieved from National Information Standards Organization website: http://www.niso.org/apps/group_public/download.php/17091/NISO RP-25-2016 Outputs of the NISO Alternative Assessment Project.pdf

Newson, R., Rychetnik, L., King, L., Milat, A., & Bauman, A. (2018). Does citation matter? Research citation in policy documents as an indicator of research impact – an Australian obesity policy case-study. *Health Research Policy and Systems*, *16*(1), 55. https://doi.org/10.1186/s12961-018-0326-9

Noyons, E., & Ràfols, I. (2018). Can bibliometrics help in assessing societal contributions of agricultural research? Exploring societal interactions across research areas. In P. Wouters, R. Costas, T. Frassen, & A. Yegros-Yegros (Eds.), *Science, Technology and Innovation indicators in transition*. Retrieved from https://hdl.handle.net/1887/64521

Okamura, K. (2019). Interdisciplinarity revisited: evidence for research impact and dynamism. *Palgrave Communications*, *5*(1), 1–9. https://doi.org/10.1057/s41599-019-0352-4

Ortega, J.-L. (2020). Altmetrics data providers: A meta-analysis review of the coverage of metrics and publication. *El Profesional de La Información*, *29*(1). https://doi.org/10.3145/epi.2020.ene.07

Porter, A. L., Garner, J. G., & Crowl, T. (2012). Research coordination networks: Evidence of the relationship between funded interdisciplinary networking and scholarly impact. *Bioscience*, *62*(3), 282–288. Retrieved from http://www.jstor.org/stable/10.1525/bio.2012.62.3.9

Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, *81*(3), 719–745. https://doi.org/10.1007/s11192-008-2197-2

Pulido, C. M., Redondo-Sama, G., Sordé-Martí, T., & Flecha, R. (2018). Social impact in social media: A new method to evaluate the social impact of research. *PLOS ONE*, *13*(8), e0203117. https://doi.org/10.1371/journal.pone.0203117

Robinson-Garcia, N., van Leeuwen, T. N., & Ràfols, I. (2018). Using altmetrics for contextualised mapping of societal impact: from hits to networks. *Science and Public Policy*. https://doi.org/10.1093/scipol/scy024

Science-Metrix. (2018). *Review of the Human Frontier Science Program 2018*. Retrieved from https://www.hfsp.org/node/12547#book/

Tahamtan, I., & Bornmann, L. (2020). Altmetrics and societal impact measurements: Match or mismatch? A literature review. *El Profesional de La Información*, *29*(1), e290102. https://doi.org/10.3145/epi.2020.ene.02

Tattersall, A., & Carroll, C. (2018). What can Altmetric.com tell us about policy citations of research? An analysis of Altmetric.com data for research articles from the University of Sheffield. *Frontiers in Research Metrics and Analytics*, *2*(9). https://doi.org/10.3389/frma.2017.00009

Thelwall, M. (2016). Web indicators for research evaluation: A practical guide. In *Synthesis Lectures on Information Concepts, Retrieval, and Services* (Vol. 8). https://doi.org/10.2200/s00733ed1v01y201609icr052

Thelwall, M. (2018). Altmetric prevalence in the social sciences, arts and humanities: Where are the online discussions? *Journal of Altmetrics*, *1*(1). https://doi.org/10.29024/joa.6

Thelwall, M., & Kousha, K. (2015). Web indicators for research evaluation. Part 1: Citations and links to academic articles from the Web. *El Profesional de La Información*, *24*(5), 587. https://doi.org/10.3145/epi.2015.sep.08

van der Hel, S. (2016). New science for global sustainability? The institutionalisation of knowledge co-production in Future Earth. *Environmental Science and Policy*, *61*, 165–175. https://doi.org/10.1016/j.envsci.2016.03.012

Wang, L., & Li, Z. (2018). Effects of knowledge flows from public science to industrial technologies. *Proceedings of the 2018 STI Conference*, 576–588. Retrieved from https://openaccess.leidenuniv.nl/bitstream/handle/1887/65342/STI2018_paper_109.pdf?sequence=1

Wang, Q., & Schneider, J. W. (2019). *Consistency and validity of interdisciplinarity measures*. https://doi.org/10.1162/qss_a_00011

Willis, C. D., Riley, B., Stockton, L., Viehbeck, S., Wutzke, S., & Frank, J. (2017). Evaluating the impact of applied prevention research centres: results from a modified Delphi approach. *Research Evaluation*, *26*(2), 78–90. https://doi.org/10.1093/reseval/rvx010

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., … Johnson, B. (2015). *The metric tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. https://doi.org/10.13140/RG.2.1.4929.1363

Wouters, P., Zahedi, Z., & Costas, R. (2018). Social media metrics for new research evaluation. In W. Glanzel, H. F. Moed, S. U., & M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators* (pp. 687–713). Retrieved from http://arxiv.org/abs/1806.10541

Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. *PLOS ONE*, *13*(5), e0197326. https://doi.org/10.1371/journal.pone.0197326

Getting in touch with the EU

IN PERSON
All over the European Union there are hundreds of Europe Direct information centres.
You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

ON THE PHONE OR BY EMAIL
Europe Direct is a service that answers your questions about the European Union.
You can contact this service:
– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696, or
– by email via: https://europa.eu/european-union/contact_en

Finding information about the EU

ONLINE
Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU PUBLICATIONS
You can download or order free and priced EU publications from:
https://op.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU LAW AND RELATED DOCUMENTS
For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

OPEN DATA FROM THE EU
The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

The Key indicators for Research and Innovation project was a three-year project designed to support the development and implementation of the Commission's R&I policies with evidence at the EU, country and regional levels. As a result, 11 publications reflecting the work of the project were produced:

- The final study report makes a summary of the work done for each of the six work packages of the project;
- The final study approach report covers the methodology applied to collect and treat the data and the computation of the indicators;
- The ten policy briefs provide an in-depth analysis of different research questions with the aim of formulating policy recommendations in support of evidence-based policymaking.

Studies and reports

Publications Office
of the European Union