Review

The economic impacts of open science:

a rapid evidence assessment

Michael J. Fell [1*]

[1]   University College London; michael.fell@ucl.ac.uk
*   Correspondence: michael.fell@ucl.ac.uk

Abstract: A common motivation for increasing open access to research findings and data is the potential to create economic benefits – but evidence is patchy and diverse. This study systematically reviewed the evidence on what kinds of economic impacts (positive and negative) open science can have, how these comes about, and how benefits could be maximized. Use of open science outputs often leaves no obvious trace, so most evidence of impacts is based on interviews, surveys, inference based on existing costs, and modelling approaches. There is indicative evidence that open access to findings/data can lead to savings in access costs, labour costs and transaction costs. There are examples of open science enabling new products, services, companies, research and collaborations. Modelling studies suggest higher returns to R&D if open access permits greater accessibility and efficiency of use of findings. Barriers include lack of skills capacity in search, interpretation and text mining, and lack of clarity around where benefits accrue. There are also contextual considerations around who benefits most from open science (e.g. sectors, small vs larger companies, types of dataset). Recommendations captured in the review include more research, monitoring and evaluation (including developing metrics), promoting benefits, capacity building and making outputs more audience-friendly.

Keywords: open science; open access; open data; economic impacts

1. Introduction

There has been a trend towards increasing openness in research practices in the United Kingdom (UK) and internationally. There is now an expectation that, except under certain circumstances, UK Research Council funded research will be published in open access format and that underlying data will be freely shared [1,2]. The European Union has an extensive programme of work on open science, including a declaration of open science principles [3] and a monitoring programme [4]. In the United States, the Fair Access to Science and Technology Research Act requires publicly funded scientific research to be made freely accessible online [5]. Open practices are expected to bring 'significant social and economic benefits' [1, p1], among other benefits [6]. However, while previous reviews have touched on this topic [7,8], no systematic attempt has yet been made to identify and synthesize evidence relating to this claim and present a clear picture of the economic impacts that open science might have, how these comes about, and how benefits might be maximized.

The aim of this rapid evidence assessment (REA) is to bring together the best available evidence to answer the question: *what are the economic impacts of open science?* It has the following specific objectives:

- To identify the types of direct and indirect economic impact that open science has been empirically demonstrated to have.
- To identify the mechanisms by which economic impacts come about.

- To identify the contextual factors that affect whether or not (and the extent to which) economic impacts occur, and the extent to which the open science approach was a necessary condition of their occurring.
- To assess the magnitude and relative importance of different types of impact (positive and negative).
- To identify methods by which economic impacts have been (or have been suggested to be) recorded and/or quantified, and how counterfactuals (i.e. non-open science approaches) have been estimated.
- To identify policies (or other interventions including [but not limited to] clearer / better communications about knowledge exchange and open science) which can help maximize the economic benefits and reduce the costs associated with open science.
- To identify trade-offs between economic and other (e.g. academic, societal) impacts.

The term 'open science' has a broad range of meanings, ranging from the publishing of research outputs (rather than keeping them confidential) to free sharing of every aspect of research including protocols, analysis plans and notebooks [9]. The scope of this rapid review if restricted to open access to research findings (usually through open access publishing) and data (open research data), since these are the most mature areas where there is likely to be the most evidence, and where funders' requirements are currently most explicit. Hereafter in this paper I use the term 'open science' to refer to these concepts – but it is important to bear in mind that the review does not capture all impacts that might be associated with employing the full suite of open science practices.

The review takes in direct economic impacts in which open science has been a contributory factor, including changes in productivity, competitiveness, employment, income, investment, and value. The focus was on impacts in the economy in general, not those restricted only to the ecosystem around scholarly communication (e.g. publishers, university library budgets, etc.), since the latter topic has been widely considered in the context of the debate around transition to open access publishing. Indirect benefits – such as increases in economic productivity through health improvements resulting from new drugs developed with the contribution of open science outputs – are recognized as potentially very important, but are out of scope. It is important to emphasize that economic impacts are certainly not the only, or necessarily the most significant, area where open science has to the potential to make a positive (or negative) contribution – access to knowledge and data has intrinsic and extrinsic benefits that go well beyond this – but these impacts are also not considered in this review.

In the next section I describe the systematized review method, and then go on to present results of the review. I break these down into sections on methods employed to assess economic impacts, the impacts themselves (broken into benefits and costs/challenges), contextual factors, and recommendations made in the reviewed material. I finally draw conclusions and make a number of recommendations.

2. Materials and Methods

An REA is a review that uses systematic review methodology to map the characteristics of what is already known about the research topic, and supports decision making by providing evidence on topics of policy interest and identify gaps in research. They can be limited in focus to fit with resource constraints and policy timescales, and are now quite widely employed in UK policcmaking (e.g. see [10,11]). The main advantages over non-systematic reviews are the level of objectivity they provide as to what should be included, and the increased confidence they give that no important evidence has been missed.

In conducting an REA there must be a trade-off between the breadth of material which it is possible to cover and the comprehensiveness of the review within given resources. Conducting a systematic and comprehensive review in a narrow area can be preferable because this allows us to say with greater confidence that no key evidence has been missed in the area of focus, and it is subsequently possible to keep this review up to date without having to revisit older material to

94 check for gaps. The scope can be expanded in future as resources become available, again without
95 having to back-track over previously reviewed material.
96   For this reason, alongside others described in the previous section others described above, this
97 focused predominantly on the economic impacts of free dissemination of research methods and
98 results (open access publishing) and of research data (open research data). Pilot searches suggest
99 that performing a comprehensive review in these subjects will be possible with the resources
100 available, but that adding other relevant subjects (such as open source software for research)
101 substantially expands the amount of material to review.   However, where relevant evidence was
102 encountered during the course of a search focused on open access/data (for example, where specific
103 open source software is required to make use of open research data), it will be recorded. For the
104 avoidance of doubt, evidence dealing with open data that is not primarily generated for research
105 purposes is out of scope.
106   REAs require a protocol to be set out in advance specifying how material will be located, what
107 criteria will be used to assess whether it will be included in the review, what information will be
108 extracted from included sources, and how this will be synthesized. The original protocol for this
109 review is available at https://osf.io/jd3eb. Administrative restrictions meant that it was not possible
110 to share this protocol publicly at the time of the initial review. However, input was sought from a
111 number of leading experts in the open science domain and their recommendations incorporated.
112   The remainder of this section describes the process that was followed, and highlights any
113 departures from the original protocol, along with a justification. I follow the PRISMA guidelines [12]
114 as closely as possible.

115 *Search, screening and quality assessment*

116   The search strategy specifies how the source material for the review was identified. This was
117 achieved through a combination of online searches, citation checking and expert consultation.
118 Online searches require the concepts of interest to be specified, and then operationalized for use in a
119 search. Table 1 shows the concepts and search terms that have been identified, and shows an
120 illustrative search string for the Scopus database.

121   Table 1: Construction of database search terms.

	Open science	**Economic impact**
Concept	Open science Open data Open research data Open access Open metrics	Economic impact Financial impact Monetary impact Cost/benefit analysis Input-output General equilibrium modelling Return on investment Growth accounting
Search term	"open scien*" "open data" "open research data" "open access" W/1 publ* OR paper* OR journal* OR book* "open metric*"	econom* financ* cost* mone* cba bca "input-output" "general equilibrium" "return on investment" "growth accounting"
Scopus example	TITLE-ABS-KEY ("open scien*" OR "open data" OR "open research data" OR ("open access" W/1 publ* OR paper* OR journal* OR book*) OR "open metric*") OR TITLE ("open	

	access") AND TITLE-ABS-KEY (econom* OR financ* OR cost* OR mone* OR cba OR bca OR "input-output" OR "general equilibrium" OR "return on investment" OR "growth accounting")

122
123   The searches are focused on title/abstract/keywords of articles to increase the proportion of hits
124 likely to be directly related to the subject in question, in the knowledge that some relevant results
125 may be missed. Pilot searches indicated that expanding searches to the full text of documents would
126 lead to an unmanageably large number of hits. The search term "open access" has been qualified in
127 the search to minimize the number of results referring to articles that themselves are open access
128 (where the words "open access" often appear in the copyright statement with the abstract). At the
129 protocol review stage a number of additional search terms were suggested, including abbreviations
130 CBA and BCA, and terms such as "general equilibrium models", "return on investment" and
131 "growth accounting". All searches were recorded and are available to view in Appendix B.
132   This search strategy was deployed on the following databases which permit Boolean searching:

133 • Scopus
134 • Web of Science (all databases)
135 • ScienceDirect

136   The following sources were also searched based on combinations of the above search terms.
137 • JISC
138 • UK Government website (gov.uk)
139 • Innovate UK
140 • UK Research Council and HEFCE websites
141 • Google scholar (searches of title only, limited to review of first 300 results [13])
142 • Open Data Institute
143 • Digital Curation Centre (DCC)
144 • Nesta
145 • Centre for Open Science
146 • Open Research Funders Group
147 • Open Scholarship Initiative
148 • Open Access Bibliography
149 • Open Access Directory
150 • Universities UK
151 • OECD Library
152 • Europa.eu
153 • European Universities Association

154   All hits from database searches were downloaded and imported into reference manager
155 software (Zotero). For searches of other sources, results were downloaded where the title and/or
156 initial screening suggests the document is likely to meet the inclusion criteria. In some cases, where
157 Google was employed to search for sources, results were restricted to PDF documents. This is
158 because pilot searches revealed that without this restriction, an unmanageably large number of
159 general (e.g. news page) results were often returned. Reports of empirical research were almost
160 always found to have PDF version available. While this approach risks missing sources, it was
161 considered justified given that other forms of source identification (such as citation checking and
162 expert elicitation) were also being employed, and given the significant time savings it allowed. All
163 documents so far identified as part of the work of the Open Science team were also be captured
164 through a review of internal documents. Two informal discussions were also conducted with
165 experts in the field, and any sources they recommend were captured[1]. As the original searches were

---

[1] These interviews will also be used to inform and refine the case study objectives and interview schedules.

166 conducted in late 2017 with an informal update in early 2018, a limited number of further update
167 searches were conducted during the final preparation of this paper in April 2019.
168   All references were imported into the systematic review management software EPPI-Reviewer
169 4 and de-duplicated. They were then screened by a single reviewer on the basis of the
170 inclusion/exclusion criteria set out in Table 2. A first-pass screening done on title/abstract, with a
171 second screening on the full document.

172   Table 2: Inclusion and exclusion criteria.

Include if source	Exclude if source
Is in English	Is not in English
Includes discussion or analysis of making research methods/findings freely available (open access publishing). AND/OR Includes discussion or analysis of making research data freely available (open research data).	Does not include discussion or analysis of the economic impacts of making research methods, findings or data openly available.
Includes explicit consideration (informed by empirical evidence) of direct or indirect economic impacts (with or without quantitative value estimates).	Does not explicitly consider economic impacts based on empirical evidence.
	Only considers open source software without discussion of open data or open access publishing.
	Reports work in an open access publication or based on open data, but is not explicitly about these concepts.
	Focuses only on impacts within the scholarly communications ecosystem of publishers, universities and research funders.[2]
	Only considers open data in general, as opposed to specifically research data.

173
174   The reference lists of all documents still included at this point were checked, and any relevant
175 documents referred to were also loaded into EPPI-Reviewer and passed through the screening
176 process. The final list of included documents was reviewed by a number of colleagues, who were
177 invited to suggest documents which may have been missed – these were screened for inclusion. Any
178 documents which came to my attention at any point were also loaded in and screened. A flow chart
179 showing this process, and the numbers of records retained/removed at each stage, is included in
180 section 3.
181   Quality assessment was conducted to ensure that poor quality studies are not given undue
182 weighting when synthesizing evidence. Because of the exploratory nature of this review, no sources
183 were completely excluded on the basis of quality. The TAPUPAS framework was employed to aid
184 assessment of quality, with sources graded on each point from poor (0 points) to good (2 points)
185 (original protocol stipulated a 1-5 point scale, but this was subsequently considered unrealistically
186 precise). It was also anticipated that for any full economic impact assessments located, the
187 EPPI-Centre data extraction form (economic) section 7 items would be used for quality assessment.
188 In the event, this more detailed approach was not feasible given project resource constraints.

189

---

2 This criterion was added after the protocol was finalised as many documents were identified in this category and, as set out in the introduction, such impacts were not the focus of this review.

190   *Extraction*

191   Data on the following criteria were extracted from all included sources:

192   •   Organizational affiliation of authors
193   o   Academia
194   o   Industry
195   o   Governmental
196   o   Other non-governmental
197   •   Study country
198   •   Funding source (if applicable)
199   •   Coverage
200   o   Open access publishing
201   o   Open research data
202   o   Other
203   •   Key aims
204   •   Data collection and analysis methods
205   o   Interviews
206   o   Social surveys
207   o   Macroeconomic modelling
208   o   Cost-benefit analysis
209   o   Case study
210   o   Other
211   •   Key findings (on economic impact)
212   •   Relevant policies cited
213   •   Recommendations
214   •   Quality assessment (scored 0-2)
215   o   Transparency
216   o   Accuracy
217   o   Purposivity
218   o   Utility
219   o   Propriety
220   o   Accessibility
221   o   Specificity

222   Additional coding of extracted information was conducted in NVivo (qualitative analysis
223   software) to allow identification of key themes.

224   *Synthesis*

225   The results were synthesized under headings aligned with the review objectives presented
226   above. For each objective a thematic analysis approach was used. For example, to describe the types
227   of economic impact identified in the review, specific examples were collected together under broad
228   common themes (such as 'labour cost savings'). Impact themes were associated with quantitative
229   data where possible (e.g. where examples of financial costs, job growth/destruction, etc., can be
230   allocated under themes). Themes were also be identified in contextual factors, and any consistent
231   associations with economic impacts highlighted (a statistical analysis was not possible).

232   *Updating*

233   The original review conducted over late 2017 and early 2018. It was therefore updated prior to
234   preparation of this paper in April 2019, by repeating a subset of the original searches, restricted by
235   date to 2018 and 2019 (see Appendix B). Documents identified during this process were screened
236   and data extracted according to the process described above.
237

238   *Description of results*

239   The following sections present the results of the review, including a summary appraisal of
240   methods used in the studies which were identified, the benefits and challenges they uncovered (with
241   value estimates where available), a consideration of contextual factors that may be relevant to the
242   results, and a summary of recommendations made in the sources to support benefits maximization.
243   In presenting this kind of synthesis it is important to bear in mind that the detailed sources that were
244   drawn on often include caveats and contextualization which it is neither possible nor desirable to
245   capture in full. I attempt to detail the main limitations or contextual points where relevant. Please
246   note that these sections also refers to additional literature, beyond that identified in the review,
247   where it can help add useful context or additional information.

248   **3. General review summary**

249   Figure 1 shows a flowchart of the REA, including number of documents included/excluded at
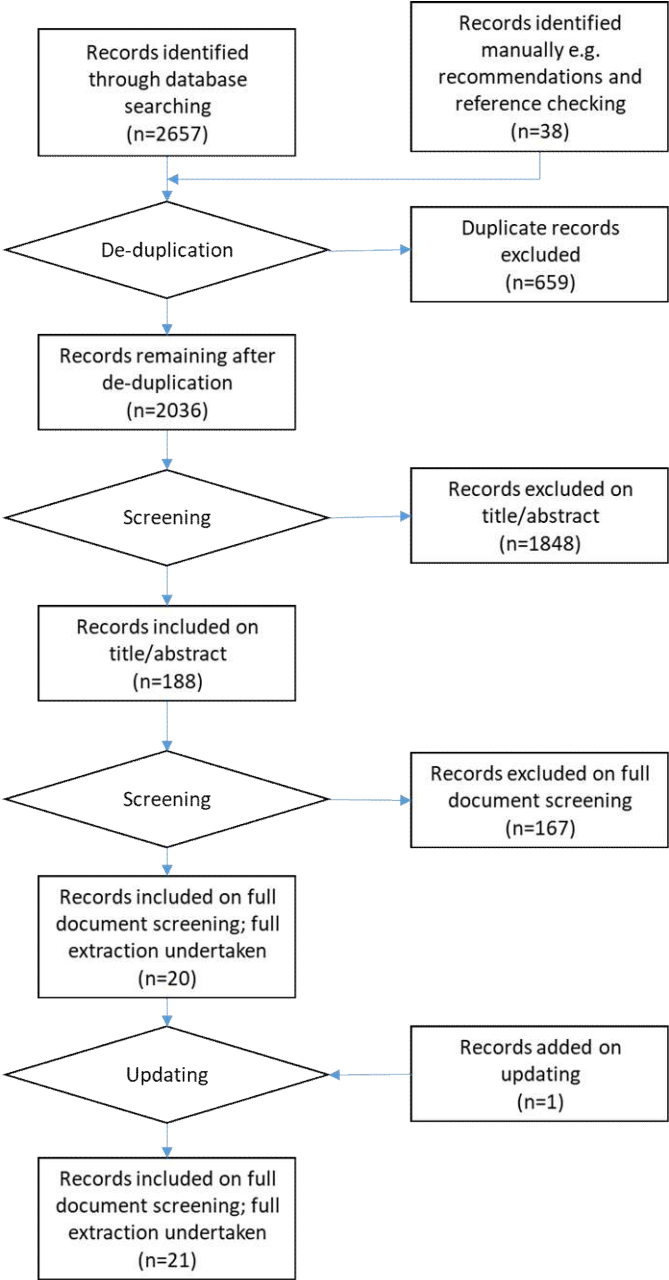250   each stage.

251

252   Figure 1: Flow chart showing REA process.

253   The original review identified 20 documents for detailed extraction, with updating providing
254   one additional document, leading to a total of 21. There was substantial variety in the type of study
255   identified, which can present challenges for quality assessment. A quite flexible QA method
256   (TAPUPAS[3]) was selected in anticipation of this, and ultimately two studies were de-prioritized on
257   the basis of insufficient methodological detail, or justification of findings/conclusions. A further five
258   documents did not present new empirical evidence but were still included because of their relevance
259   to the review questions (for example one study presented a possible research method).

260   Five sources described studies focusing specifically on the UK, while the remainder focused on
261   the USA, Canada, Denmark, Finland, or more than one country. All but seven of the sources were
262   published in or after 2014. There was a roughly even balance in coverage between open access
263   publishing (eight sources) and research data (nine sources), with six sources also including
264   consideration of other open science topics. As indicated in the previous section, the review identified
265   a lot of work has been done on the economics of open access publishing, but the vast majority of this
266   focuses on the impacts within the university/publisher/funder ecosystem rather than the broader
267   economy, and was therefore not included.

268   There was a high concentration of included studies around certain researchers or research
269   approaches. John Houghton, for example, was involved as co-author in six, and many of these
270   studies used similar research approaches. Furthermore, because these studies draw on more diverse
271   evidence sources in comparison to some other included documents (often, for example, combining
272   interviews, surveys and desk research to inform economic analysis, rather than presenting only
273   interview evidence), and more explicitly focus on questions of economic value of open science, they
274   are drawn upon comparatively more heavily in this review. A table summarizing the main aims and
275   methods of each included study is included in Appendix A.

276   **4. Methods for assessing economic impacts**

277   The review uncovered a diverse range of approaches to gauging the economic impacts of open
278   science. This is likely related to some degree to the difficulty in tracking the usage of open science
279   outputs outside of the academic context (where citation practices make this more achievable). Often
280   users leave no footprint when they access open science outputs, and it there is often no requirement
281   to declare usage. It is possible (although there is no evidence to support this) that firms may perceive
282   it to be beneficial to conceal their use of open findings and data if they believe they derive an
283   advantage from using resources or approaches that competitors may not be aware of. Even if usage
284   is apparent, it is still often difficult to attribute specific economic impacts to the openness of the
285   outputs.

286   This section sets out the range of ways in which researchers have attempted to combat these
287   challenges and provide value estimates. Principally these involved using surveys/interviews to elicit
288   user information and opinion, and cost-benefit analysis, sometimes informed by modelling
289   approaches (based on desk research on costs/benefits, user input and assumptions). These may be
290   conducted either in the context of a specific case (such as a research collaboration of data repository)
291   or applied more generally (such as at country level). Often several approaches are combined in the
292   same study. The methodological heterogeneity makes it harder to draw comparisons between
293   findings of separate studies than if standard approaches were employed.

294   *Question-based approaches*

295   Where it is difficult to measure revealed behaviour (such as observing actual use of open
296   science outputs), it is common to ask users about their usage instead. In their evaluations of UK data
297   repositories [14] and the European Bioinformatics Institute [15] (along with evaluations in other

---

3

https://www.scie.org.uk/publications/researchmindedness/findingresources/assessingresearchquality/

298 countries not included here), Beagrie and Houghton used surveys and interviews for this purpose.
299 Interviews were used to help develop the survey content, such as by suggesting which kinds of
300 cost-benefit to include questions about. They also permit more in-depth and detailed insight into
301 how costs or benefits might come about. Interviews are usually only possible with relatively small,
302 non-random samples of participants, and are therefore not suitable for making generalizations to
303 wider populations on, for example, proportions of users affected by certain issues. Interviews were
304 employed in a number of other studies in the review [16–20].
305   The Beagrie and Houghton surveys asked users to make estimates of factors such as the time
306 spent preparing data for upload, time spent working with data they had downloaded, and the extent
307 to which they would be able to perform their work without access to the open data. Answers to such
308 questions, combined with estimate of salary costs, permit insight into investment in preparing/using
309 open data (known as Investment Value). Among other questions, the surveys also asked users to
310 estimate how much they would be willing to pay to access the (currently free) services, or be paid to
311 forfeit access (willingness to accept). These Contingent Valuation approaches permit estimation of
312 the value placed on free services – with willingness to accept potentially giving insight into value to
313 users whose budgets would prevent them suggesting a high willingness to pay. Subtracting
314 willingness to pay from use value (in a welfare approach) gives insight into consumer surplus. The
315 main drawback with surveys is that it is hard to know how far responses reflect 'objective reality'.
316 For example, in the Beagrie and Houghton work discussed above, respondents were asked to assess
317 how far having access to data in a repository affected their work efficiency (on a percentage scale).
318 Even when pilot testing has been done, it is unlikely that all respondents would interpret such a
319 question in the same way, or be able to come up with an accurate estimate (or even to know exactly
320 what efficiency would mean in this context). The authors are keen to highlight such limitations, but
321 it is easy for such caveats to get lost when key findings are extracted from their context. Surveys
322 were also employed by other studies in the review [17,20].

Cost-benefit analysis (and similar) approaches

324   In the studies reviewed, data from surveys and interviews was often combined with
325 information and assumptions based on desk research to inform variants of cost-benefit analysis
326 (CBA). In a standard CBA researchers sum up the identifiable existing or expected costs and benefits
327 (converted to monetary terms) to which stakeholders in the system of interest are exposed to
328 calculate the overall costs and benefits which accrue, and to whom. For example, in an evaluation of
329 the Human Genome Project, Battelle were able to calculate the value of additional employment
330 provided by the collaboration and organizations connected with it, including the tax revenue this
331 yielded [21]. Assessments of the value of open access publishing such as those by Houghton et al.
332 [22] and CEPA LLP and Mark Ware Consulting Ltd. [23] use best estimates of costs in the existing
333 system and then use modelling approaches to suggest how these might change under alternative
334 publishing models. In their study of the value of text mining (including consideration of the benefits
335 of open access to this), McDonald and Kelly [16] drew on CBA techniques recommended in HM
336 Treasury's Green and Magenta Books (on appraisal and evaluation in central government), and
337 Orange Book (on the management of risk).
338   This CBA-like approach such as that used in CEPA LLP and Mark Ware Consulting Ltd. [23]
339 sheds useful light on costs/benefits within the scholarly communication ecosystem, but not the
340 wider impacts on the economy. Multiple works by Houghton and others (for example see [22])
341 therefore include an additional modelling step. They employ a modified Solow-Swan model, which
342 is a way of estimating the contribution of technological progress (driven by R&D) to economic
343 growth. It can be used to estimate a return on investment in R&D through value to the economy in
344 the coming years. In the sources reviewed, the model is modified to include consideration of the
345 accessibility of R&D outputs (i.e. the extent to which they are available to inform technological
346 progress) and efficiency with which R&D expenditure is converted into technological development
347 (which could be potentially be improved by open science practices which help reduce redundancy
348 or improve replicability – see next sub-section). The modified model does not tell us anything about

the real extent to which open science approaches could improve accessibility of efficiency (these are based on assumptions), but does provide an indication into what different levels of accessibility/efficiency improvement might mean for value in years to come.

While studies of open access publishing tend to focusing on estimating costs/benefits across the sector, much of the empirical evidence on economic impacts of open science more broadly is collected in the context of case studies. For example, this includes studies of research collaborations such as the European Bioinformatics Institute [15], the Structural Genomics Consortium [24], the Research Collaboratory for Structural Bioinformatics Protein Data Bank [25] or of data repositories such as the Economic and Social Data Service [26]. As well as making estimates of the general value of these initiatives using the techniques described above, such evaluations frequently include sub-case studies. In the case of the EBI, for example, part of the work involved conducting impact case studies on specific companies which have been involved in the initiative. By pulling out findings from the research for specific named companies, the research reveals concrete examples of how impact comes about. While the generalizability of such examples is limited, they do serve to illustrate when specific routes to impact are working, or to highlight routes which may otherwise not have been anticipated.

Other approaches employed in evaluation of impacts of open data in general have looked even more broadly at its possible impacts, including consideration of Social Return on Investment (proposed approach only) [27] and wide ranging economic analysis (by McKinsey) of value to sectors such as education, healthcare and oil/gas [28]. This analysis looks at issues such as, in the transportation sector, improved infrastructure planning and management, optimized fleet investment and better-informed consumer decision-making. Specific details of the McKinsey calculations are not provided, but they results in very high value estimates ($3 trillion per year to the global economy). Nevertheless another open data report for the G20 considers that the McKinsey estimates could even prove to be underestimates [29]. It is important to note that open data in general tends to differ from open research data in ways that are relevant to economic impacts – see subsection on contextual considerations, below.

5. Economic impacts

The review identified a variety of ways in which economic costs and benefits associated with open access to research findings and data, and a number of value estimates. These are set out in the following section. The benefits part is separated into two broad routes – *efficiency* and *enablement*.

- Efficiency means getting the same output from research or innovation for less input (principally public research funding). For example, if open access publishing can be shown to allow access to research findings for the same number of researchers for a lower overall cost, this would represent an efficiency saving. While this review is not principally concerned with economic impacts within the university/publishing sector, such efficiencies are also relevant to wider economic impacts through improved returns to R&D.
- Enablement signifies ways in which open science approaches have led to economically impactful activities which would have been less likely to occur in a closed environment.

The remainder of this section presents evidence on efficiency and enablement benefits, costs and challenges, contextual issues and the recommendations made by the authors of the documents reviewed.

Efficiency

The main identified efficiency benefits come about in the form of access costs savings, potential for labour cost savings (or productivity improvements), and other benefits such as reduced transaction costs. These sub-section explores evidence for these benefits in more detail.

Access cost savings

Under a closed model, if a user wants to have access to certain findings or data they must either pay for that access, or pay to create the findings/data themselves. There is suggestive evidence of savings potential associated with both open to research findings and data. Such savings would accrue both to public funders of research (as researchers are large consumers of research outputs) and commercial users.

Research for Jisc in 2011 [30] found that more than two-thirds (68%) of researchers in UK universities and colleges felt they did not have access to a sufficiently wide range of journals and conference papers; only 18% of researchers in industry/commerce felt themselves to have similarly limited access. However, for those participants who viewed having access to journal papers as important, a quarter of those in industry/commerce described their current level of access as fairly/very difficult, compared to just 5% in universities and colleges. The main barrier to access was unwillingness to pay. Research on access to research outputs in Denmark [17] found examples of companies resorting to workarounds such as establishing formal/informal relationship with university researchers and students to gain access vicariously, or physically going to university library buildings to use their terminals for access.

Houghton and others [22] summed the costs and savings associated with accessing research findings (e.g. production, publishing, dissemination) under open or paid access models. According to their analysis, open access approaches would have been £813-1180 per journal article cheaper than toll access, equating to £80-116m per year (in 2007) for the UK higher education sector, and with similar potential levels of saving relating to book publishing. These figures are disputed, however – especially by those analysts who question the accuracy of the publishing cost estimates [31,32]. This and other analysis [23] have informed the UK's current policy position which is to move towards open access publishing models. Recent research for Universities UK suggests that the costs of transitioning to open access are being impacted by increases in Article Processing Charges [33], which might be expected to negatively impact the net benefits.

Research data, unlike publications, has not tended to be accessed by organizations on a subscriptions basis, so a similar analysis is not really useful in this area. However, recent years have seen a move towards the development of data repositories, where datasets can be accessed for free by most users with no or minimal registration requirements. Beagrie and Houghton conducted studies in an attempt to evaluate the savings that arise to users through having access to such repositories. These include the Economic and Social Data Service (ESDS, now integrated into the UK Data Service) [26], the Archaeology Data Service (ADS) [34], the British Atmospheric Data Service (BADS, now part of the CEDA Archive) [35] and the data repositories curated by the European Bioinformatics Institute (EBI) [15]. It is important to note that some of these repositories do not only include open research data, but also a substantial quantity of open government data. The values included below do not distinguish between the different sources of data, so will overestimate the value connected with (solely) research data. Access to most UK Data Service data is only free for research and teaching purposes; commercial users are required to pay.

The authors used surveys to ask users (mostly from the research sector) of these services what they would pay to retain access (willingness to pay), and what payment they would accept for loss of access (willingness to accept). They also asked users to estimate how much time they spent using the service (and preparing data to submit to it), to which it was possible to apply a labour cost. As highlighted in the section of this paper considering methods, getting an accurate estimate of willingness to pay/accept from surveys can be difficult for reasons including respondents' own difficulties in estimating value, incentives to obscure true value estimates, and mismatch between valuations and real purchasing behaviour [36].

These caveats accepted, in the case of the ESDS they found a consumer surplus of £21m per year – that is, users expressed willingness to pay £21m more than the £4m they were currently investing in labour costs to use the service. Given the cost of running the service of roughly £3m per year, this equates to a net economic value of £18m per year, or more than five times the cost of running the service. A similar study on the European Bioinformatics Institute suggests that users' willingness to pay, at around £322m, exceeded the annual operational budget of around £47m by approximately

448 six times. A study which employed similar methods to estimate the economic impact of the Research
449 Collaboratory for Structural Bioinformatics Protein Data Bank also found large positive returns – for
450 example, contingent value (or hypothetical willingness to pay multiplied by number of users
451 estimated on the basis of website visits) was more than 100 times greater than the project's annual
452 operating cost [25]. It should, however, be noted that this study relies extensively on assumptions
453 based on previous valuation work.

454 Measuring users' willingness to pay can underestimate the value they attach to a service where
455 those users are quite resource constrained. This was the case for the ADS, where users' willingness
456 to pay for the service was sufficiently low that the consumer surplus and net economic value were
457 negative. The authors argue that in some cases (for example in sectors where funding is relatively
458 scarce) it is more instructive to focus on users' willingness to accept payment for lack of access,
459 which has no resource constraints. If this is used instead of willingness to pay, a net economic value
460 for the service of roughly eight times the £700k operational budget is achieved. The authors note a
461 similar resource constraint for users of the BADC, although in this case their stated willingness to
462 pay still suggests a positive net economic value of just under £1m, on an operational budget of just
463 over £2m. These specific examples illustrate the challenges in providing anything more than very
464 general indications of the potential economic impact of free resources, and provide (at best) limited
465 evidence for benefits potential.

466 As well as reducing direct costs associated with access to research outputs, open science
467 approaches can help reduce transactions costs. The Structural Genomics Consortium is a
468 collaborative research body that aims to support drug discovery by sharing open access information
469 on pre-competitive biological structures[4]. An evaluation of its work found that its open collaborative
470 approach avoided the direct and labour costs of establishing multiple material transfer agreement
471 between partners [24]. Another study has estimated that the costs of such agreements for a single
472 collaboration can run into the hundreds of thousands of dollars [37], so given the extensive
473 collaboration (between hundreds of facilities) that the SGC has enabled, this element of saving is
474 likely to be considerable.

475 Transaction cost savings are also likely to arise for text/data mining uses. Under normal
476 circumstances, those intending to text/data mine are required to reach agreement for such usage
477 with multiple publishers and content providers [16]. The introduction of a text and data mining
478 copyright exception has reduced this burden for non-commercial uses, but commercial use is still
479 restricted and there are other barriers related to mass downloading of material and sharing of
480 outputs [38]. Campaigns are underway at the EU level to address such issues [39]. As well as the
481 costs and time required to reach such agreements, it also introduces significant uncertainty into such
482 projects as it is possible that some agreements may not be reached. Open access approaches have the
483 potential to mitigate such costs[5].

484 Labour cost savings or productivity improvements

485 It takes time for firms and researchers to actually access research outputs, and this time is
486 reflected in labour costs. It can take longer for people to access closed research outputs than when
487 access is open. A survey in Denmark by Houghton, Swan and Brown found that knowledge-based
488 SME employees spent on average 51 minutes to access the last research article they had difficulty
489 accessing, and this rose to 63 minutes for university researchers [17]. Given they reported difficulty
490 accessing 17 articles per year, the cost of such delays based on average staff salaries could amount to
491 €70m (although the authors note that their sample was not representative)[6]. This is a cost that could,
492 in theory, be saved if all access barriers were removed. However, the actual time saving associated

---

4 https://www.thesgc.org/about/what_is_the_sgc

5 Please note, however, that no financial estimates were discovered for the scale of such savings.

6 For comparison, UK GDP is approximately 8.5 times that of Denmark.

493 with open access was not tested. Support was also found for this potential saving in an
494 interview-based study by Parsons et al. [18].
495   The data repository evaluations by Beagrie and Houghton referred to above also attempted to
496 estimate the time savings that accrued to users of the repositories. This was done by asking them to
497 estimate how many hours they spent working with repository data (to which a value could be
498 attached based on average salaries), and to use a percentage scale to estimate how much more
499 efficiently this allowed them to use their time than if the repository were not available. This is clearly
500 a difficult assessment to make, and the researchers report problems associated with participants
501 interpreting the questions in different ways. However, open-ended response fields allowed some
502 insight into this interpretation – for example users taking into account reduction in form-filling,
503 traveling to on-site/supervised repositories, and efficiencies associated with the format of the data.
504 Using these responses, the researchers suggest there may be efficiency gains worth £68-100+ million
505 per years for the ESDS, and smaller but also positive savings for ADS (£13-58m) and BADC
506 (£10-58m)[7]. They estimate the equivalent savings for the EBI at approximately £1-5 billion, at least 20
507 times more than the direct operational cost.
508   As well as making individual research outputs easier to access, open access can also facilitate
509 text and data mining, which allows generation of new information through analysis of large bodies
510 of text or data. One of the key benefits of this is that time taken to extract useful information from
511 sources can be reduced (compared to manual approaches). It is difficult to estimate what the level of
512 time saving might be, but McDonald and Kelly suggest that [16]:

> *If text mining enabled just a 2% increase in productivity corresponding to only 45 minutes per academic*
> *per working week, this would imply over 4.7 million working hours and additional productivity worth between*
> *£123.5m and £156.8m in working time per year.*

516   It should be noted that some applications of text and data mining, such as filtering and
517 presenting only relevant research to users, risk negative consequences such as missing key evidence
518 unless   a substantial proportion of available content is available for mining.

519 Other efficiency benefits

520   There are a number of other ways in which open science could lead to more efficient research,
521 and therefore more positive economic impact. It is estimated that closed research can lead to high
522 levels of duplication – that is, where separate teams work on the same thing unbeknownst to each
523 other. While the closed nature of the research inherently makes the extent of this hard to estimate, an
524 analysis of pharmaceutical patents by 18 large companies showed that 86% of target compounds
525 were investigated by two or more companies [40]. It is important to draw a distinction between
526 duplication and replication (the latter of which can help increase confidence in conclusions);
527 duplication is a problem because multiple companies may invest in developing a compound (or
528 other innovation) that has already been demonstrated to be ineffective [41]. Redundancy in research
529 goes wider than questions of open science and is connected with challenges such as general
530 under-publication of 'null' results [42]. However, open approaches such as pre-registration of trials
531 could help ameliorate the situation, in this case by helping to reduce suppression of 'null' results and
532 therefore potential for wasteful duplication of approaches which have been shown to be ineffective.
533 Text and data mining could also make it easier to identify previous findings [16]. The challenge of
534 decreasing research productivity has been highlighted [43]; the above factors have the potential to
535 mitigate this.
536   So far the evidence has focused on efficiency, or doing the same with less input. The implication
537 here, for example, is that firms are able to access research findings and data in closed situations, but
538 just at higher financial and other resource costs. However, there are also mechanisms by which an
539 open approach can enable access, connections and collaborations that would not have happened, or
540 would have happened more slowly, otherwise. This is the subject of the next section.

---

[7] Please note that the authors emphasize that, because of some differences in the way these figures
were calculated between studies, they are not directly comparable with each other.

Enablement

Enablement benefits come in the form of new products, services, companies and collaborations, as well as permitting work that could not otherwise have been undertaken.

New products/services/companies

It has been suggested that open science approaches have the potential to lead to the development of new products, services, companies and even industries [29,44]. This review did not reveal any economy- or even industry-wide attempts to estimate the extent to which this has taken place. However, it did identify a number of case studies that demonstrate the existence of such mechanisms and provide some insight on the circumstances under which they come about. These examples are all drawn from the life sciences (see section on contextual issues below for further discussion of this).

Probably the largest project to make research data openly available was that Human Genome Project, which ran from 1990 to 2003 and was funded principally by the US government (at $3.8 billion). Data were made publicly available within 24 hours of discovery [45]. The availability of the human genome has led to the development of new diagnostic tests and technologies, and informed many new disease treatments and treatment approaches [45]. For example, use of genetic information has informed new tests which allow anticipation of the likelihood of adverse reactions – this reduces harm to vulnerable patients, while allowing effective drugs to continue to be used in people to whom they do not pose a risk [21]. Further products and services have been enabled in a range of domains, from agriculture and environment to forensics and industrial biotechnology. An economic impact assessment of the project found that it generated 'an economic (output) impact of $796 billion, personal income exceeding $244 billion, and 3.8 million job-years of employment' (pES-2) [21].

Because that report does not focus on the 'openness' of the project in particular, it does not permit assessment of how the benefits would have compared under a more closed model. However, other research has taken advantage of the fact that a parallel, comparatively closed sequencing effort (run by the firm Celera) yielded data with some IP protection [46]. It estimates that 'Celera's IP led to reductions in subsequent scientific research and product development on the order of 20–30 percent' (p24), a result of the additional transaction costs that using Celera data would impose. This finding is caveated that this may not represent research effort that was 'lost' – it may simply have been focused on non-Celera genes.

There are examples of new products and companies emerging from other, similar projects. The Structural Genomics Consortium is a not-for-profit public private partnership which conducts research on protein structures, releasing structural data freely and without patent protection [47]. An evaluation found that about half of a small sample of researchers surveyed (17 respondents) believed that their research would lead to development or trialling of a pharmaceutical product in future, and three people reported that their research already had this outcome [24]. A good example is JQ1, a compound that research through the Consortium suggested may be used to inform development of cancer treatments. Following this discovery, GlaxoSmithKline started its own proprietary research programme which led to a clinical trial within two years [48]. Research that later compared the number of patents connected to the JQ1 discovery found that after four years, 105 patents had been filed, compared to a mean of 29 for two similar compounds which were not released openly [49]. Three spin-out companies have resulted from the work of the SGC, with one, Tensha Therapeutics, connected with JQ1 [48]. It received $15m initial investment, and was acquired within a year (by Roche) for $535m [50].

Collaborations

As before, it is not always clear that new products or companies would not have appeared had the originating project not been open in nature. However, the experience of the SGC points to an additional benefit that such open models can yield that means that research can happen that may not

590 have done otherwise – that of new pre-competitive multi-stakeholder collaborative research. A case
591 study of the consortium [51] highlighted the potential conflicts that can arise in standard
592 university-firm partnerships, such as firms' desire to appropriate knowledge outputs running
593 counter to academics' requirement to publish openly, or for universities to make their own IP claims.
594 The SGC managed this in a variety of ways, such as by revealing lists of target proteins without
595 attaching companies' identities to them (so protecting companies' research interests), and not
596 revealing the list of targets publicly. Instead, protein structures were released only when resolved.
597 By providing such assurance – identified as 'meditated revealing' and 'enabling multiple goals' – the
598 SGC was able to attract funding from the Wellcome Trust, GlaxoSmithKline, Novartis, Merck,
599 government organizations and other smaller foundations. In 2011 the consortium reported the
600 addition of two new members, including Pfizer, and that together the consortium members were
601 committing $50m for the next four years of research along with $9m in-kind contributions [52].
602 Another major open collaboration is in its early stages at the Montreal Neurological Institute and
603 Hospital [53]. While no evaluation is yet available of its impacts, an evaluation plan is in place which
604 will also look more broadly at impacts of open science approaches internationally, potential for new
605 metrics, and other relevant questions [54].

606   Open science collaborations are also appearing in other subject areas, such as the SPOMAN
607 (Smart Polymers and Nano-composites) project at Aarhus University [55]. In this case the focus is
608 more on collaborative ideation, with partner companies being involved in determining research
609 priorities. The research is focused on fundamental knowledge and partners may not patent it.
610 Instead, all outputs (included data, code and lab notebooks) are shared openly, with companies
611 being free to file patents on specific applications of the findings. No economic evaluation is yet
612 available on the project.

613 Permitting work

614   A key enabling contribution of open science outputs can be to permit further research that
615 would not otherwise have been possible. For example, the evaluation of the European
616 Bioinformatics Institute found that 45% of survey respondents could not have either found the data
617 they access through the repository themselves, not created it themselves [15]. Work based on this
618 data can therefore be viewed as additional to what would have occurred had the Institute not existed
619 (although any such counterfactual must take into consideration how funding for the Institute might
620 otherwise have been deployed). As similar proportion of survey respondents to the Economic and
621 Social Data Service evaluation reported this also (330 out of 894, 37%) [26].

622   It has also been suggested that enhanced text- and data-mining capabilities permitted by open
623 approaches (as outlined above) can result in previously hidden connections between different areas
624 of research being unveiled [16]. The example is provided of a tool which was used to analyse disjoint
625 biomedical literatures to identify possible new disease targets for the drug thalidomide, which could
626 then be subject to further research which might not otherwise have occurred [56]. As highlighted
627 above, if R&D is expected to have positive growth impacts in the economy, then any research which
628 is permitted by open science outputs and which is genuinely additional (for example, as compared
629 to duplicating existing data) contributes to this.

630   As discussed in the 'methods' section above, macroeconomic modelling approaches have been
631 used to explore the potential impact of more open science approaches on the returns to R&D in the
632 wider economy. Specifically, they look at the impact of varying the following factors [57]:

633 •   Accessibility, or the extent to which research findings can be accessed by users.
634 •   Efficiency, or the extent to which R&D generates knowledge that is useful socially or
635   economically.

636   The variables summarize many of the impacts described above. Such modelling requires many
637 large assumptions so the results are naturally highly caveated. However, it has been estimated that a
638 conservative 5% increase in accessibility and efficiency in the UK could have been worth £172m per
639 year in increases returns to public sector R&D [22]. Because such returns are experienced year after

640 year (assuming the accessibility/efficiency gains are permanent), this can contribute to increased
641 growth.

Costs and challenges

643   The review also set out to see if there was evidence of open sciences approaches having
644 negative economic impacts. While many examples of specific additional or different costs were
645 identified (such as extra costs of preparing datasets for publication or payment of Article Processing
646 Charges), none of the overall value estimates located were negative. Going beyond direct costs, a
647 number of ways were proposed in which either the full benefits of open science might not be
648 realised, or there may even be some negative effects. However, none of the work identified
649 attempted to put a value on these effects. This section summarises the evidence underpinning these
650 concerns.
651   There were a few suggestions that a lack of capacity within firms (and elsewhere) to make use
652 of outputs from open research diminished their potential to have benefits. For example, work by
653 Houghton, Swan and Brown, in Denmark found that the second most cited difficulty in accessing
654 findings was simply searching for an article online, but not being able to find it [17]. The authors
655 suggest that this points to a lack of higher-level information literacy skills amongst SMEs. Johnson et
656 al. [58] point to the potential for similar skills shortages to lead to lack of the accountability that open
657 data (in general) is purported to bring, and similarly the lack of awareness of the value of text
658 mining is highlighted as a reason for the lack of full benefits realisation via that route [16]. Huber et
659 al. [59] also highlight awareness and skills barriers to maximizing benefits from open data (not just
660 research data) – but also highlight further barriers including uncertainty around future availability
661 of data, the risk of imitation (as others also have access to the data), and legal and reputational risks.
662   Important questions are also raised about where the benefits of participating in open science
663 accrue, and what this might mean for the conduct of research in general. Part of this is directly
664 related to the rewards of commercialization. For example, while early stage research collaborations
665 such as the SGC have proved successful, later stage applied research (with more immediate
666 expectation of benefits) has been less amenable to open approaches due to firms' perceived
667 difficulties in protecting intellectual property at this stage [41]. Morgan argues that if governments
668 can provide 'regulatory exclusivity' to drugs developed in an open way, this could provide an
669 alternative form of protection [60]. There is also the question of indirect benefits or disadvantages to
670 researchers of conducting open research. The potential conflict is highlighted of institutional
671 expectations to publish in certain journals, which may not provide open access [61]. Researchers
672 may also be concerned about lack of credit for producing open science outputs [62].
673   Some work has focused on the potential incentive for researchers to delay publication of papers
674 because such publication would come with a requirement to share their data, potentially
675 diminishing their opportunity fully exploit it [63]. The requirement to share data could even act as a
676 disincentive to collect it (because the researcher may expect to be able to reap the full benefits of their
677 data collection work). The research, a modelling study which examined researcher incentives to
678 generate and share data, found that if mandatory data disclosure requirements led to strategic
679 delays in publication, this could affect the overall (economic) welfare of the research community.
680   There has been additional speculation that researchers may perceive a tension between their
681 institutions' or funders' expectations to commercialize their work, and the desire or requirement to
682 share findings and data openly. While in principle such a tension can be shown to exist (in the policy
683 statements of individual funders regarding commercialization and openness), there is as yet no
684 empirical evidence that researchers recognize it [64]. This review did not reveal any cost-benefit
685 analyses of the value to the economy of direct commercialization of publicly funded research (e.g. by
686 university patents or spinouts) compared to commercial exploitation of open science outputs.
687 Finally, in the context of open data in general, the issue was raised of whether public funding for
688 data which is likely to be of commercial use represents a subsidy [58]. This could be a concern if the
689 commercial offerings informed by the data had limited benefits for citizens.

690   *Contextual issues*

691   The review revealed concentrations of open science approaches in some research areas. All the
692   economic evaluations of public/private open research collaborations focused on the life sciences –
693   such as the HGP, SGC and EBI. The common factor was a desire amongst collaborators to jointly
694   support (rather than duplicate) basic analysis which could subsequently be used to inform product
695   development. As Savage points out, efforts to extend such collaboration further into the drug
696   discovery process have met with less success [41]. Economic evaluations were not found for
697   collaborations adopting open approaches in other areas. The reasons for this are not clear – it may be
698   that decreasing rates of translation of basic discoveries into commercial treatments, combined with
699   the expiry of existing drug patents, has led pharmaceutical companies in particular to embrace open
700   innovation approaches [65].   The sector is also subject to strong regulation which may incentivise
701   cost sharing more than in other sectors. However,   there is no reason to think that life sciences is
702   unique in offering the possibility for collaboration on basic research to underpin future
703   developments. There is the example of SPOMAN [55] in materials science (as described briefly
704   above), and a potential parallel in the example of the National Geological Repository (NGR),
705   operated by the publicly funded British Geological Survey. A Natural Environment Research
706   Council impact study highlighted the role of core samples made openly available by the NGR in
707   increasing the efficiency of drilling and geotechnical companies' operations.

708   There is some evidence to suggest that economic impacts may be more like to accrue to certain
709   types of commercial actor. A survey by Ware [66] found that 86% of large firm respondents reported
710   accessing research information through company subscriptions compared to 77% of small and
711   medium sized enterprises (SMEs), while use of open access sources was 68% in large firms
712   compared to 71% in SMEs. On a monthly basis, 27% of large firms respondents used an in-house
713   information service compared to 15% of SMEs. While Ware does not explicitly make this point, it
714   therefore seems likely that SMEs would stand to gain disproportionately from an increase in open
715   access to research findings, although this does not take into account factors such as capacity to make
716   use of such information (see recommendations section below).

717   Based on wider (informal) reading of the literature on open data business models, it is apparent
718   that many rely on access to real-time data such as on meteorological or transport conditions (e.g.
719   CityMapper[8]). While there are examples of real-time open research data feeds (such as LondonAir[9],
720   which provides information on levels of air pollution), the examples identified in the review were
721   exclusively concerned with discrete, fixed research datasets. Even in cases where data is released on
722   a frequent basis (such as in the Human Genome Project, which released data within 24 hours of
723   collection), this does not permit CityMapper-like business models meaning this avenue for
724   commercialization is likely closed off for the majority of open research data. The case has also been
725   made that the granularity of research data reduces its commercial potential. Tim Vines, writing on
726   the Scholarly Kitchen website, argues that unlike social media or city data which have individuals or
727   locations as a common thread, most scientific datasets are 'small and collected to answer a very
728   specific question' [67].

729   In some of the sources reviewed there is discussion of localisation (to the country of origin) of
730   returns to R&D and, by inference, open access to research findings/data. For example, Houghton and
731   others [22] draw on previous suggestions that domestic knowledge should be weighted at 66-73%
732   compared to 27-33% for foreign knowledge in terms of value to the domestic economy. However,
733   they do not make an estimate for whether/how opening access to findings and data might affect
734   these weightings. At the very local level, no evidence was found regarding economic impacts in the
735   direct locality of research teams producing open outputs. However, as mentioned in the 'Efficiency'
736   section above, firms do sometimes benefit from proximity to a local university by accessing research
737   information in their library. An increase in open accessibility of findings/data may make such
738   proximity 'workarounds' less important.

---

8   https://citymapper.com/london

9   https://www.londonair.org.uk/LondonAir/Default.aspx

Recommendations captured in the review

Many of the documents included in the review included recommendations related to increasing positive economic impacts of open science and support further study. The majority of the recommendations are based in general promotion of open science approaches. This might include, for example: ensuring that funding is available for data preparation or publishing fees [22]; increasing provision of institutional or subject repositories [22]; creating more data journals [63]; and encouraging universities to support immediate data disclosure [63]. There was also a call for special access conditions to research outputs for text and data mining [16]. There were some recommendations to focus attention on awareness of the benefits of use of open science outputs [19]. Parsons et al. found that lack of search skills and interpretive expertise was highlighted as a barrier to use of open access resources [18]. They recommend that making repositories more user-friendly (such as by simplifying user interfaces, providing lay summaries, improving business-relevant metadata and advising academics on how to present findings for business audiences).

Specifically in relation to economic impacts, the most common call was for more research, monitoring and evaluation in this area, aided by identification of new metrics. As previous sections have highlighted, there is a lack of available, comparable data on actual use of open science outputs in a way that would allow identification of economic impacts. Research has tended to rely on either reported use (e.g. through surveys), assumptions, or detailed case studies. Conventional research indicators such as citation counts have little relevance to questions of economic impact, while innovation indicators such as patents are either less relevant (in the context of open innovation) or yield little insight on the contribution of open science outputs to the innovation. Without strong evidence for ways in which positive economic impacts of open science can be maximized, it is difficult to design effective support policies.

Reflecting on their evaluations of data repositories [14] and the EBI [15], the authors recommend that such bodies collect/report more data on costs, usage and users (voluntarily), and that data be collected in standardized ways allowing more comparative and granular analysis. Initiatives are underway that could support this, such as Crossref's DOI Event Tracker pilot [68]. Houghton and Sheehan [44] called in 2006 for development of better ways of tracking the translation of knowledge generated by R&D into economic value (and how this depends on openness of outputs), and identifying the extent to which value is localized or spills over to other localities. While there has indeed been progress in research on these areas (e.g. [69,70]), the review did not identify evidence on economic impacts of open science incorporating these developments (although research is underway at McGill university in Canada that will include consideration of local impacts [54]). Chataway et al. [61] point out the need for a better understanding of how open access to research outputs influences firms' use of research, especially in the case of research data. They highlight efforts that are underway to support this [71]. Houghton et al. make the case for research metrics and incentives that better reflect the value of innovative open models of research communication [22]. Parsons et al. [18] provide a package of research recommendations including on how open access status affects search engine discoverability, awareness and use of open materials in different organizational contexts, usability of repositories to business users and the role of different organizational/capacities in enabling use of open materials.

4. Conclusions and recommendations

Conclusions

In this paper, I have presented the results of a Rapid Evidence Assessment which set out to identify that best available evidence of the economic impacts of open science. This section briefly summarizes the key findings and considers their possible implications for policymakers.

Use of open science outputs (e.g. by firms) often leaves no obvious trace, so most evidence of impacts is based on interviews, surveys, inference based on existing costs, and modelling approaches. For example, surveys have been used to ask open data users how much time they spend searching for data and how this is reduced when the data is open, then using salary estimates to

assess labour cost savings. Willingness to pay/accept approaches have also been used to estimate the value of free resources. Such approaches introduce significant potential for measurement error, and assumptions on existing costs (such as those involved in publishing findings) are often disputed.

The review identified evidence that open access to findings and data can lead to positive economic impacts through (a) efficiency and (b) enablement, but the evidence base patchy and diverse. Efficiency means getting the same output from research or innovation for less input (principally public research funding). For example, if open access publishing can be shown to allow access to research findings for the same number of researchers for a lower overall cost, this would represent an efficiency saving. Enablement signifies ways in which open science approaches have led to economically impactful activities which would have been less likely to occur in a more closed environment.

Regarding efficiency, there is indicative evidence that open access to findings/data can lead to savings in access costs, labour costs and transaction costs. Open access should lead to cost-savings in the university/publishing ecosystem, although (as highlighted above) modelling assumptions are disputed and there are significant costs in transitioning to some models. Open access could reduce costs to firms of accessing research findings (or increase availability), although this may only be a 'latent' savings if firms don't have capacity to access/use findings/data. Easier access to research findings/data can reduce time required to access them, saving labour costs. There is some evidence of reduction in transaction costs, such as material transfer agreements or for text mining (although open access in itself does not eliminate such barriers except for non-commercial uses, where a text and data mining copyright exception applies).

There are examples of open science enabling new products, services, companies, research and collaborations. Multi-stakeholder collaborations such as the Human Genome Project, Structural Genomics Consortium and European Bioinformatics Institute were enabled by agreements around early, open release of data. The SGC revealed a compound (JQ1) which had cancer treatment potential. GSK has gone to clinical trials with products based on this. Research that later compared the number of patents connected to the JQ1 discovery found that after four years, 105 patents had been filed, compared to a mean of 29 for two similar compounds which were not released openly. Three spin-out companies have resulted from the work of the SGC, with one, Tensha Therapeutics, connected with JQ1 . It received $15m initial investment, and was acquired within a year (by Roche) for $535m. Evaluation of the European Bioinformatics Institute found that 45% of survey respondents could not have either found the data they access through the repository elsewhere, nor created it themselves.

Modelling studies indicate higher returns to R&D with the greater accessibility and efficiency that open access allows. Macroeconomic modelling approaches have been used to explore the potential impact of more open science approaches on the returns to R&D in the wider economy. Specifically, they look at the impact of varying the following factors:

- Accessibility, or the extent to which research findings can be accessed by users.
- Efficiency, or the extent to which R&D generates knowledge that is useful socially or economically.

It was estimated in 2009 that a conservative 5% increase in accessibility and efficiency in the UK could have been worth £172m per year in increases returns to public sector R&D. Such models include many general assumptions.

While the review identified specific additional or different costs associated with greater openness (such as extra costs of preparing datasets for publication), none of the overall net value estimates located were negative. Key barriers to use of open science outputs included lack of skills capacity in search, interpretation and text mining, and lack of clarity around where benefits accrue. There is evidence for lack of higher-level information skills amongst SMEs. Modelling work has suggested that mandatory data sharing could cause researchers to delay publication, affecting

839 overall welfare of research community. There are suggestions (but little evidence) of tension
840 between pressures to commercialise research and a more open approach.
841   There are also contextual considerations around who benefits most from open science (e.g.
842 sectors, SMEs vs larger companies, types of dataset (e.g. real-time vs static), extent of local benefits.
843 All the major research collaborations evaluated were in the life sciences, with collaborators jointly
844 supporting basic science. Later stage research collaborations have met with less success. SMEs are
845 less likely to have institutional subscriptions (e.g. to journals) than larger companies so could benefit
846 more from free access, but they may also be subject to more time and skills constraints. Research
847 data tends to be more static and granular than other forms of open data (e.g. public transport
848 departures), reducing its commercial potential. There is as yet no evidence on the role of open
849 science in localization of benefits, although previous general estimates have suggested that domestic
850 knowledge has more value locally than foreign knowledge.
851   Recommendations include more research, monitoring and evaluation (incl. metrics), promoting
852 benefits, capacity building and making outputs more audience-friendly. There is a lack of available,
853 comparable data on actual use of open science outputs – where possible, repositories and open
854 collaborations should collect more data on costs, usage and users. A toolkit for tracking open science
855 impacts has recently been developed to help improve and formalize this process [72]. Research is
856 needed on how openness affects local vs global return to R&D. Better understanding is needed of
857 how firms make use of R&D outputs, including open outputs. Consideration should be given to
858 promoting the benefits of open material and building capacity within firms to locate and use it.
859 Repository design and content could be improved to make access and use easier for a wide range of
860 users (such as by simplifying user interfaces, providing lay summaries, improving business-relevant
861 metadata and advising academics on how to present findings for business audiences).

862 *Recommendations*

863   The above findings demonstrate the potential of open science to provide a range of economic
864 benefits. On the basis of these findings, I make the following recommendations for policymakers and
865 research funders. These intentionally do not include measures to either increase the quantity of open
866 outputs available, or the usability/interoperability/etc. of these outputs – these are goals of broader
867 open science policy. Instead, they focus on specific ways in which the positive economic impacts of
868 open science might be maximized. The relevance of each recommendation will vary depending on
869 country context.
870   **Promote and support new open collaborations.** This review provides a number of examples of
871 multi-stakeholder public-private collaboration to tackle pre-competitive research challenges. These
872 were shown to support increases in R&D funding and development of new products and
873 companies. Research funders should review their areas of responsibility for subjects meeting
874 appropriate characteristics (i.e. basic research likely to inform subsequent innovation, multiple large
875 private competitors, significant regulatory burden) and promote collaborations in areas revealed as
876 suitable. General principles and guidance should be developed to balance a requirement for
877 openness with corporate partners' commercial considerations (as for the Structural Genomics
878 Consortium Open Science Trust Agreement). The benefits of such collaborations should be fully
879 explored and actively promoted to potential commercial partners.
880   **Streamline text and data mining.** At the moment there is an array of licence conditions even
881 around many open access outputs that represent barriers to text and data mining, either forbidding
882 it or putting up excessive transaction costs. In the UK, a text and data mining (TDM) copyright
883 exception has gone some way to addressing this, but still precludes commercial applications and
884 presents other barriers. This is also a live issue at the time of writing as debates continue around the
885 introduction of the new EU Directive on Copyright in the Digital Single Market. Licensing should be
886 further simplified with a presumption that text and data mining are possible in a wider range of
887 applications (campaigns are underway at the EU level to support this). Infrastructure should be put
888 in place to support TDM access to open access material for firms and academic libraries. The
889 EU-funded OpenMinTeD project is a possible example of such infrastructure [73].

Develop an open access findings/data portal targeted at business. Most infrastructure for the outputs of research is primarily targeted at the research community, which may diminish its usability and apparent usefulness to many companies. Dedicated portals should be developed (informed by user research and building on lessons from previous similar initiatives) which provides a more relevant front-end architecture, for example categorizing open outputs by relevant business sector or business function, and prioritizing lay summary information where available. To minimize resource requirements this could take the form of an overlay on existing repository infrastructure. A curation service could allow datasets or streams which appear most likely to be useful to business (determined by characteristics such as ability to link to other datasets, real-time, etc.) to be foregrounded or even actively pushed to suitable businesses. Such portals would also act as a focal point for promoting the business benefits of open science outputs and providing support and training material for those businesses seeking to extract most value from such outputs. Models for this exist in the form of services such as Konfer [74] and the new LORIC centre at Bishop Grosseteste University [75], and universities could have a key role to play as neutral, trusted convenors.

Produce aligned positions and guidance on open science and commercialization. There are potential contradictions in the aims of policies intended to encourage open science and those promoting commercialization of research. While there is currently little evidence on the extent to which institutions and researchers perceive contradictions, to avoid confusion there should be consistency both in the word and intention of policy. Countries should review policies and measures intended to promote commercialization and open science. Based on analysis of areas of complementarity and contradiction and existing empirical evidence on how the two can most effectively be balanced, a position and guidance should be developed which address the priorities and concerns of individual researchers, universities, firms and the wider economy.

Continue support for open research data repositories. There is good evidence that they can support positive economic outcomes (although this varies from case to case). Evidencing effectiveness and impacts would be easier if more information were collected on users and uses to which data are put (especially commercial uses which currently leave few traces). This would also allow for better international comparison.

Continue research into new metrics and incentives. Work is ongoing to develop metrics which can more usefully capture the wider impacts of open science outputs. This could also include consideration of social metrics – for example ways in which companies could indicate their use and/or approval of a dataset (similar to a 'like' on Facebook). This would act as a non-scholarly equivalent to the citation, and could provide a similar basis for research impact assessment.

Appendix A

Summary of main aims and methods of the included studies.

Ref	Study aims and approach
[14]	**Beagrie, N. & Houghton, J. The Value and Impact of Data Sharing and Curation: A synthesis of three recent studies of UK research data centres. (Jisc, 2014).** A synthesis of three similar studies, the full report for each of which was also drawn on in

	this review [26,34,35]. They are treated together here as they formed part of a programme of work with very similar aims and methods. Aims: to identify the value for, and impacts on, users and depositors of three research data repositories: the Economic and Social Data Service, the Archaeology Data Service, and the British Atmospheric Data Centre. Methods: • Semi-structured interviews with a cross-section of stakeholders at each data service (n=13 to 25). • Online surveys with users (n=299 to 141) and depositors (n=42 to 193). • Macro-economic modelling based on Solow-Swan model exploring impact of increasing accessibility and efficiency of research findings on return to research and development. • Calculation of investment value, contingent value (based on willingness-to-pay/accept), use value, welfare approaches estimating consumer surplus and net economic value, and an activity costing approach to estimate efficiency savings.
[15]	**Beagrie, N. & Houghton, J. The value and impact of the European Bioinformatics Institute. (EMBL-EBI, 2016).** Aims: To explore the costs and cost savings of the EMBL-EBI, its value to users, and wider impacts. Methods: • Interviews with 29 staff and external users. • Online survey to a convenience sample of users, 4185 usable responses (response rate 17%). • Macro-economic modelling based on Solow-Swan model exploring impact of increasing accessibility and efficiency of research findings on return to research and development. • Calculation of investment value, contingent value (based on willingness-to-pay/accept), use value, welfare approaches estimating consumer surplus and net economic value, and an activity costing approach to estimate efficiency savings. • Three detailed impact case studies.
[23]	**CEPA LLP & Mark Ware Consulting Ltd. Heading for the open road: costs and benefits of transitions in scholarly communications. (Research Information Network (RIN), JISC, Research Libraries UK (RLUK), the Publishing Research Consortium (PRC) and the Wellcome Trust, 2011).** This work is primarily focused on informing transitions to open access. Aims: Provide evidence to inform scholarly communications stakeholders around dynamics of transition towards improved access to research outputs. Methods: Regarding impacts beyond the scholarly communications system, the Solow-Swan model approach is used to quantify potential UK economy-wide impacts under different scenarios.
[61]	**Chataway, J., Parks, S. & Smith, E. How Will Open Science Impact on University-Industry Collaborations? Foresight and STI Governance 11, 44–53 (2017).** Aims: To consider the possible impacts of open science approaches on university-industry collaboration. Methods: This was not an empirical study, but combines findings from previous research to develop and support arguments.
[76]	**Giovani, B. Open Data for Research and Strategic Monitoring in the Pharmaceutical and Biotech Industry. Data Science Journal 16, (2017).** Aims: to explore how to extract value from data in the biotech sector, and how companies manage intellectual property to so as to benefit from open data while also protecting their business. Method: Interviews and an online survey. Response rates were very small and limited

	methodological detail are supplied.
[77]	**Houghton, J. Open Access: What are the Economic Benefits? A Comparison of the United Kingdom, Netherlands and Denmark. (Social Science Research Network, 2009).** Applied the same approach as [22], but also encompassing the Netherland and Denmark.
[44]	**Houghton, J. & Sheehan, P. The economic impact of enhanced access to research findings. (Centre for Strategic Economic Studies, Victoria University, 2006).** Aims: To assess the value of increasing open access to research findings in OECD countries. Methods: • Macro-economic modelling based on Solow-Swan model exploring impact of increasing accessibility and efficiency of research findings on return to research and development.
[17]	**Houghton, J., Swan, A. & Brown, S. Access to research and technical information in Denmark. (2011).** Aims: to examine access to, and use of, technical information by knowledge-based SMEs in Denmark, as well identifying barriers, costs and benefits. Approach: • 23 interviews with representatives of SMEs, some of which participated in a national incubator scheme, some of which had not (follow-up interviews were also conducted based on survey responses. • Online survey sent to around 1000 knowledge-based SMEs (non-random), 98 usable responses.
[22]	**Houghton, J. et al. Economic implications of alternative scholarly publishing models: Exploring the costs and benefits. (Jisc, 2009).** Aims: To identify costs and benefits of open access within scholarly publishing, as well as to the UK economy more broadly. Methods (relating to broader economic impacts outside the scholarly communications ecosystem, which are the subject of this review): • Macro-economic modelling based on Solow-Swan model exploring impact of increasing accessibility and efficiency of research findings on return to research and development.
[58]	**Johnson, P. A., Sieber, R., Scassa, T., Stephens, M. & Robinson, P. The Cost(s) of Geospatial Open Data. Transactions in GIS 21, 434–445 (2017).** This source does not focus on research data specifically, but was included as it provides a useful perspective on costs associated with open data. Aims: To identify externalized and unintentional impacts of open data. Methods: This was not an empirical study, but combines findings from previous research to develop and support arguments.
[24]	**Jones, M. M. et al. The Structural Genomics Consortium: A Knowledge Platform for Drug Discovery. 19 (RAND Corporation, 2014).** Aims: To understand the nature and diversity of benefits gained through the Structural Genomics Consortium both for partners and the wider research community including consideration of relative merits of the open vs more closed models of operation. Methods: • Document and literature review. • Semi-structured interviews with researchers/collaborators (18), current of former funders (17), and external stakeholders (9). • Online survey of SGC principal investigative researchers. • Economic impact assessment of SGC outputs. • An internal future scenarios workshop.
[29]	**Lateral Economics. Open for Business: How Open Data Can Help Achieve the G20 Growth Target. (Omidyar Network, 2014).** Aims: Quantify and illustrate potential of open data (in general, but including research data)

	to meet G20's growth target (with focus on Australia), and to estimate what proportion of the target open data policy could deliver. Methods: • Macro-economic modelling based on Solow-Swan model (for Australia) exploring impact of increasing accessibility and efficiency of research findings on return to research and development. • Seven descriptive case studies illustrating how various impacts came about.
[16]	**McDonald, D. & Kelly, U. Value and benefits of text mining. (Jisc, 2017).** Aims: To explore value and benefits of text mining under existing and alternative licensing conditions, including consideration of costs, risks, and barriers. Methods: • Targeted consultation with key stakeholders. • Desk research. • Cost benefit analysis based on economic analysis of all evidence identified, considering cost savings and productivity gains, potential for wider impact/innovation, efficiency and fairness of the market. • Case studies based in part on existing research. • Also drew on baseline evidence from other countries.
[63]	**Mueller-Langer, F. & Andreoli-Versbach, P. Open access to research data: Strategic delay and the ambiguous welfare effects of mandatory data disclosure. Information Economics and Policy 42, 20–34 (2018).** Aims: To investigate the effects of mandatory data disclosure policies on researcher decisions (e.g. around publication and data disclosure). Methods: Mathematical modelling of researcher decisions.
[20]	**ODI. Open data means business. (Open Data Institute, 2015).** Aims: To identify use of open data by UK companies (without specific focus on research data, but use of scientific and research data is considered). Methods: • Online survey with 79 responses (response rate of approximately 20%). • Follow-up interviews with 12 companies who completed the surveys.
[18]	**Parsons, D., Willis, D. & Holland, J. Benefits to the private sector of open access to higher education and scholarly research. (Jisc, 2011).** Aims: To identify and quantify benefits to the private sector of open access to university research outputs, along with the enablers of these benefits, mechanisms and contextual factors. Methods: • Systematic literature review. • Interviews with business bodies (14), semi-structured telephone interviews with 44 enterprises, and nine detailed enterprise profiles also based around interviews.
[48]	**RAND Europe. Open Science Monitoring Impact Case Study – Structural Genomics Consortium. (European Commission Directorate-General for Research and Innovation, 2017).** Aims: To give an overview of the impact of the Structural Genomics Consortium. Methods: This was not an empirical study, but combines findings from previous research to describe impacts (including economic/financial).
[25]	**Sullivan, K. P., Brennan-Tonetta, P. & Marxen, L. J. Economic Impacts of the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank. RCSB Protein Data Bank (2017). doi:10.2210/rcsb_pdb/pdb-econ-imp-2017** Aims: To examine the value and economic impacts of the RCSB. Methods: This study largely followed the research approach set out in [25]. Information on (for example) willingness to pay and salary costs were transferred directly from that study, while usage data were based on usage of the RCSB website.

[8]	Tennant, J. P. et al. The academic, economic and societal impacts of Open Access: an evidence-based review. F1000Research 5, 632 (2016).
	Aims: To review evidence on academic, economic and social impacts of open access publishing, including brief consideration of open research data.
	Methods: This was not an empirical study, but gives examples of impacts based on a non-systematic literature review. A short section is included considering evidence on economic impacts on non-publishers.
[21]	Tripp, S. & Grueber, M. Economic Impact of the Human Genome Project. (Battelle Memorial Institute, 2011).
	Aims: To assess the economic (and other) impacts of genome sequencing.
	Methods: Input-output modelling based around direct investment in the HGP, investments in follow-on research connected with the HGP, and the wider genomics industry developed and fostered through the HGP.
[19]	Tuomi, L. Impact of the Finnish Open Science and Research Initiative (ATT). (Profitmakers Ltd., 2016).
	Aims: To analyse the impacts of the Open Science and Research Initiative both nationally and internationally, and to offer recommendations for remainder of the programme.
	Methods:
	• Interviews with industry representatives.
	• Email questionnaires, 'web brainstorming' and desk research.

935 **Appendix B**

936   The following table records the search strings used to identify material in online searches, the
937 dates of searches, the number of records collected in each search, along with explanatory comments.

Database (+#)	Date	String	Hits	Comments
Scopus 1	20/09/17	TITLE-ABS ("open scien*" OR "open data" OR "open research data" OR ("open access" W/1 publ* OR paper* OR journal* OR book*) OR "open metric*") OR TITLE ("open access") AND TITLE-ABS-KEY (econom* OR financ* OR cost* OR mone*)	1926	Removed keywords from first part as some papers with open data tag that fact there.
Web of Science 1	20/09/17	(TS=("open scien*" OR "open data" OR "open research data" OR "open access publ*" OR "open access paper*" OR "open access journal*" OR "open access book*" OR "open metric*") OR TI=("open access")) AND TS=(econom* OR financ* OR cost* OR mone*) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.	982	
Science Direct 1	20/09/17	(tak("open scien*" OR "open data" OR "open research data" OR "open access publ*" OR "open access paper*" OR "open access journal*" OR "open access book*" OR "open metric*") OR ttl("open access")) AND tak(econom* OR financ* OR cost* OR mone*)	197	Unclear why this is so limited compared to Scopus.

Database (+#)	Date	String	Hits	Comments
JISC 1 (via Google)	20/09/17	"open science" OR "open access" OR "open data" OR "open research data" site:jisc.ac.uk filetype:pdf	60 (2 saved)	Limited to PDF, downloaded sources judged to have a chance of passing screening.
Gov.uk (via Google)	20/09/17	"open science" OR "open access" OR "open data" OR "open research data" "research" -"open -access -land" site:gov.uk filetype:pdf	33100 (15 saved)	Limited to PDF, downloaded sources judged to have a chance of passing screening. Excluding departmental OD strategies except BIS. Reviewed until 5 pages passed with no relevant material.
Innovate UK	20/09/17	N/A – hosted at gov.uk so would be picked up by above.		
AHRC 1	20/09/17	"open science" OR "open access" OR "open data" OR "open research data" site:ahrc.ac.uk filetype:pdf	51 (7 saved)	Limited to PDF, downloaded sources judged to have a chance of passing screening.
BBSRC 1	20/09/17	"open science" OR "open access" OR "open data" OR "open research data" site:bbsrc.ac.uk filetype:pdf	60 (3 saved)	Limited to PDF, downloaded sources judged to have a chance of passing screening.
ESRC 1	20/09/17	"open science" OR "open access" OR "open data" OR "open research data" site:esrc.ac.uk filetype:pdf	78 (1 saved)	Limited to PDF, downloaded sources judged to have a chance of passing screening.
EPSRC 1	20/09/17	"open science" OR "open access" OR "open data" OR "open research data" site:epsrc.ac.uk filetype:pdf	74 (5 saved)	Limited to PDF, downloaded sources judged to have a chance of passing screening.
MRC 1	20/09/17	"open science" OR "open access" OR "open data" OR "open research data" site:mrc.ac.uk filetype:pdf	107 (1 saved)	Limited to PDF, downloaded sources judged to have a chance

Database (+#)	Date	String	Hits	Comments
				of passing screening.
NERC 1	20/09/17	"open science" OR "open data" OR "open research data" site:nerc.ac.uk filetype:pdf	124 (8 saved)	Limited to PDF, downloaded sources judged to have a chance of passing screening. NOTE "open access" removed for this search as many open access papers included and over 2000 hits.
STFC 1	20/09/17	"open science" OR "open access" OR "open data" OR "open research data" site:stfc.ac.uk filetype:pdf	65 (2 saved)	Limited to PDF, downloaded sources judged to have a chance of passing screening.
HEFCE 1	21/9/17	"open science" OR "open access" OR "open data" OR "open research data" site:hefce.ac.uk filetype:pdf	10 (0 saved)	Limited to PDF, downloaded sources judged to have a chance of passing screening.
Google Scholar 1	21/9/17	allintitle: economic "open science" OR "open access" OR "open data" OR "open research data"	199 (7 saved)	Downloaded sources judged to have a chance of passing screening.
Google Scholar 2	21/9/17	allintitle: impact "open science" OR "open data" OR "open research data"	90 (5 saved)	Downloaded sources judged to have a chance of passing screening. Removed 'open access' as many articles on citation impact.
ODI 1	21/9/17	"open science" OR "open access" OR "open data" OR "open research data" site:theodi.org filetype:pdf	44 (0 saved)	Downloaded sources judged to have a chance of passing screening.
ODI 2	21/9/17	"open research data" site:theodi.org	4 (0 saved)	
DCC 1	21/9/17	impact "open science" OR "open access" OR "open data" OR "open	2480 (0 saved)	First 100 hits reviewed.

Database (+#)	Date	String	Hits	Comments
		research data" site:dcc.ac.uk		
Nesta 1	21/9/17	"open science" OR "open access" OR "open data" OR "open research data" site:nesta.org.uk filetype:pdf	143 (2 saved)	Limited to PDF, downloaded sources judged to have a chance of passing screening.
Nesta 2	21/9/17	impacts "open science" OR "open access" OR "open data" OR "open research data" site:nesta.org.uk	85 (1 saved)	Downloaded sources judged to have a chance of passing screening.
Centre for Open Science 1	21/9/17	"open science" OR "open access" OR "open data" OR "open research data" site:cos.io	120 (0 saved)	
Open Research Funders Group 1	21/9/17	"open science" OR "open access" OR "open data" OR "open research data" site:orfg.org	5 (0 saved)	
Open Access Bibliography 1	21/9/17			Scanned through bibliography, more on impacts on publishers/ institutions.
Universities UK 1	27/9/17	"open science" OR "open access" OR "open data" OR "open research data" site:universitiesuk.ac.uk filetype:pdf	38 (1 saved)	Limited to PDF, downloaded sources judged to have a chance of passing screening.
Universities UK 2	27/9/17	impacts "open science" OR "open access" OR "open data" OR "open research data" site:universitiesuk.ac.uk	29 (0 saved)	
OECD Library 1	27/9/17	impacts "open science" OR "open access" OR "open data" OR "open research data" site:oecd-ilibrary.org	46 (4 saved)	Downloaded sources judged to have a chance of passing screening.
OECD Library 2	27/9/17	"open research data" OR "open data" OR "open access" in title/abstract.	24 (2 saved)	Search on OECD Library advanced search.
Europa.eu 1	27/9/17	"economic impacts" "open science" OR "open access" OR "open data" OR "open research data" site:europa.eu	2670 (7 saved)	First 50 results reviewed, progressed to next search.
Europa.eu 2	27/9/17	impacts "open science" OR "open access" OR "open data" OR "open research data" site:europa.eu	49000 (12 saved)	First 100 results reviewed.

Database (+#)	Date	String	Hits	Comments
Europa.eu 3	27/9/17	"open research data" site:europa.eu	2800 (1 saved)	Link saved is to EU Open Research Data Pilot – to look at in general.
EUA 1	27/9/17	"open science" OR "open access" OR "open data" OR "open research data" site:eua.be	514 (0 saved)	More on implications for universities.
Google 1	27/9/17	"economic impacts" "open science" OR "open access" OR "open research data" filetype:pdf	81300 (6 saved)	First 80 results reviewed (until no relevant links for several pages).
Google 2	27/9/17	"economic impacts" "open science" OR "open access" OR "open research data"	184k (7 saved)	First 80 results reviewed (until no relevant links for several pages). Did not download some docs which were also identified in other searches.
Google 3	27/9/17	economic impacts "open science" OR "open access" OR "open research data"	59m	No additional useful docs identified in first 50 results.
Scopus 2	27/9/17	TITLE-ABS ("open scien*" OR "open data" OR "open research data" OR ("open access" W/1 publ* OR paper* OR journal* OR book*) OR "open metric*") OR TITLE ("open access") AND TITLE-ABS-KEY (cba OR bca OR "input-output" OR "general equilibrium" OR "return on investment" OR "growth accounting")	21	Terms suggested by reviewer.
Scopus 3	5/12/17	ALL("open access" W/2 "research data")	90	Added to include this term.
Scopus 4	13/4/19	TITLE-ABS-KEY ("open scien*" OR "open data" OR "open research data" OR ("open access" W/1 publ* OR paper* OR journal* OR book*) OR "open metric*") OR TITLE ("open access") AND TITLE-ABS-KEY (econom* OR financ* OR cost* OR mone* OR cba OR bca OR "input-output" OR "general equilibrium" OR	628	Update search. Items were screened by title and abstract in Scopus, and 11 which met initial criteria were downloaded.

Database (+#)	Date	String	Hits	Comments
		"return on investment" OR "growth accounting") AND (LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018))		
Google 4	13/4/19	"economic impacts" "open science" OR "open research data" filetype:pdf	No value provided (3 saved)	Update search. Limited to 2018 onwards. First 80 results reviewed (until no relevant links for several pages).
Google 5	13/4/19	"economic impacts" "open science" OR "open research data"	No value provided (0 saved)	Update search. Limited to 2018 onwards. First 80 results reviewed.

938

References

940   1.   Research Councils UK RCUK Policy on Open Access and Supporting Guidance.

941   2.   Higher Education Funding Council for England; Research Councils UK; Universities
942   UK; Wellcome Concordat on Open Research Data 2016.

943   3.   European Open Science Cloud EOSC Declaration 2017.

944   4.   European   Commission   Open   science   monitor   Available   online:
945   https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovati
946   on-policy/open-science/open-science-monitor_en (accessed on Apr 12, 2019).

947   5.   Doyle, M.F. H.R.3427 - 115th Congress (2017-2018): Fair Access to Science and
948   Technology   Research   Act   of   2017   Available   online:
949   https://www.congress.gov/bill/115th-congress/house-bill/3427 (accessed on Apr 14,
950   2019).

951   6.   Elsabry, E. Claims about benefits of open access to society (Beyond Academia).; 2017;
952   pp. 34–43.

953   7.   ElSabry, E. Who needs access to research? Exploring the societal impact of open
954   access. *Revue française des sciences de l'information et de la communication* **2017**.

955   8.   Tennant, J.P.; Waldner, F.; Jacques, D.C.; Masuzzo, P.; Collister, L.B.; Hartgerink,
956   Chris.H.J. The academic, economic and societal impacts of Open Access: an
957   evidence-based review. *F1000Research* **2016**, *5*, 632.

958   9.   OECD *Making Open Science a Reality*; Organisation for Economic Co-operation and
959   Development: Paris, 2015;

960   10. Wilson, S.; Sonderegger, S. *Understanding the Behavioural Drivers of Organisational
961   Decision-Making: Rapid Evidence Assessment*; HM Government: London, UK, 2016;

962   11. RAND Europe *What Works in Changing Energy-Using Behaviours in the Home? A
963   Rapid Evidence Assessment*; Department of Energy and Climate Change: London, UK,
964   2012;

965   12. Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis, J.P.A.;
966   Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA statement for
967   reporting systematic reviews and meta-analyses of studies that evaluate healthcare
968   interventions: explanation and elaboration. *BMJ* **2009**, *339*, b2700.
969   13. Haddaway, N.R.; Collins, A.M.; Coughlin, D.; Kirk, S. The Role of Google Scholar in
970   Evidence Reviews and Its Applicability to Grey Literature Searching. *PLOS ONE* **2015**,
971   *10*, e0138237.
972   14. Beagrie, N.; Houghton, J. *The Value and Impact of Data Sharing and Curation: A
973   synthesis of three recent studies of UK research data centres*; Jisc: London, UK, 2014;
974   15. Beagrie, N.; Houghton, J. *The value and impact of the European Bioinformatics
975   Institute*; EMBL-EBI, 2016;
976   16. McDonald, D.; Kelly, U. *Value and benefits of text mining*; Jisc: London, UK, 2017;
977   17. Houghton, J.; Swan, A.; Brown, S. *Access to research and technical information in
978   Denmark*; 2011;
979   18. Parsons, D.; Willis, D.; Holland, J. *Benefits to the private sector of open access to
980   higher education and scholarly research*; Jisc: London, UK, 2011;
981   19. Tuomi, L. Impact of the Finnish Open Science and Research Initiative (ATT). **2016**.
982   20. ODI *Open data means business*; Open Data Institute, 2015;
983   21. Tripp, S.; Grueber, M. *Economic Impact of the Human Genome Project*; Battelle
984   Memorial Institute, 2011;
985   22. Houghton, J.; Rasmussen, B.; Sheehan, P.; Oppenheim, C.; Morris, A.; Creaser, C.;
986   Greenwood, H.; Summers, M.; Gourlay, A. *Economic implications of alternative
987   scholarly publishing models: Exploring the costs and benefits*; Jisc: London, UK, 2009;
988   23. CEPA LLP; Mark Ware Consulting Ltd. *Heading for the open road: costs and benefits
989   of transitions in scholarly communications*; Research Information Network (RIN),
990   JISC, Research Libraries UK (RLUK), the Publishing Research Consortium (PRC) and
991   the Wellcome Trust, 2011;
992   24. Jones, M.M.; Castle-Clarke, S.; Brooker, D.; Nason, E.; Huzair, F.; Chataway, J. *The
993   Structural Genomics Consortium: A Knowledge Platform for Drug Discovery*; RAND
994   Corporation, 2014; p. 19;.
995   25. Sullivan, K.P.; Brennan-Tonetta, P.; Marxen, L.J. Economic Impacts of the Research
996   Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank. *RCSB Protein
997   Data Bank* **2017**.
998   26. Charlies Beagrie Ltd.; Centre for Strategic Economic Studies *Economic Impact
999   Evaluation of the Economic and Social Data Service*; Economic and Social Research
1000   Council, 2012;
1001   27. Stuermer, M.; Dapp, M.M. Measuring the promise of open data: Development of the
1002   impact monitoring framework.; 2016; pp. 197–203.
1003   28. Manyika, J.; Chui, M.; Groves, P.; Farrell, D.; van Kuiken, S.; Almasi Doshi, E. *Open
1004   data: Unlocking innovation and performance with liquid information*; McKinsey, 2013;
1005   29. Lateral Economics *Open for Business: How Open Data Can Help Achieve the G20
1006   Growth Target*; Omidyar Network, 2014;

1007   30.   Research Information Network; Publishing Research Consortium; JISC *Access to*
1008   *scholarly content: gaps and barriers*; 2011;

1009   31.   Davis, P. Challenging Assumptions on Open Access Cost Savings Available online:
1010   https://scholarlykitchen.sspnet.org/2009/07/16/challenging-assumptions-on-open-acce
1011   ss-cost-savings/ (accessed on Apr 12, 2019).

1012   32.   McCulloch, A. Discussion on JISC report on economic implications of alternative
1013   business models 2009.

1014   33.   Jubb, M.; Plume, A.; Oeben, S.; Brammer, L.; Johnson, R.; Butun, C.; Pinfield, S.
1015   *Monitoring the Transition to Open Access*; Universities UK: London, UK, 2017;

1016   34.   Beagrie, N.; Houghton, J. *THE VALUE AND IMPACT OF THE ARCHAEOLOGY*
1017   *DATA SERVICE: A study and methods for enhancing sustainability*; 2013;

1018   35.   Beagrie, N.; Houghton, J. *THE VALUE AND IMPACT OF THE BRITISH*
1019   *ATMOSPHERIC DATA CENTRE*; 2013;

1020   36.   Breidert, C.; Hahsler, M.; Reutterer, T. *A Review of Methods for Measuring*
1021   *Willingness-to-Pay*; 2006;

1022   37.   Lee, W.H. Open access target validation is a more efficient way to accelerate drug
1023   discovery. *PLoS Biology* **2015**, *13*, 1–9.

1024   38.   JISC The text and data mining copyright exception: benefits and implications for UK
1025   higher   education   Available   online:
1026   https://www.jisc.ac.uk/guides/text-and-data-mining-copyright-exception (accessed on
1027   Apr 12, 2019).

1028   39.   LIBER Copyright Reform: Help Us Ensure an Effective TDM Exception! *LIBER* 2017.

1029   40.   Leeson, P.D.; St-Gallay, S.A. The influence of the "organizational factor" on compound
1030   quality in drug discovery. *Nature Reviews Drug Discovery* **2011**, *10*, 749–765.

1031   41.   Savage, N. Competition: Unlikely partnerships. *Nature* **2016**, *533*, S56–S58.

1032   42.   Chalmers, I.; Glasziou, P. Avoidable waste in the production and reporting of research
1033   evidence. *The Lancet* **2009**, *374*, 86–89.

1034   43.   Bloom, N.; Jones, C.I.; Van Reenen, J.; Webb, M. *Are Ideas Getting Harder to Find?*;
1035   National Bureau of Economic Research, 2017;

1036   44.   Houghton, J.; Sheehan, P. *The economic impact of enhanced access to research*
1037   *findings*; Centre for Strategic Economic Studies, Victoria University, 2006;

1038   45.   SPARC From Ideas to Industries: Human Genome Project Available online:
1039   https://sparcopen.org/impact-story/human-genome-project/ (accessed on Apr 12,
1040   2019).

1041   46.   Williams, H.L. Intellectual Property Rights and Innovation: Evidence from the Human
1042   Genome. *Journal of Political Economy* **2013**, *121*, 27.

1043   47.   Structural Genomics Consortium Mission and Philosophy Available online:
1044   https://www.thesgc.org/about/what_is_the_sgc (accessed on Apr 12, 2019).

1045   48.   RAND Europe *Open Science Monitoring Impact Case Study – Structural Genomics*
1046   *Consortium*; European Commission Directorate-General for Research and Innovation,
1047   2017;

1048   49. Arshad, Z.; Smith, J.; Roberts, M.; Lee, W.H.; Davies, B.; Bure, K.; Hollander, G.A.;
1049   Dopson, S.; Bountra, C.; Brindley, D. Open access could transform drug discovery: A
1050   case study of JQ1. *Expert Opinion on Drug Discovery* **2016**, *11*, 321–332.
1051   50. Tensha Therapeutics Tensha Therapeutics To Be Acquired By Roche Available online:
1052   https://www.businesswire.com/news/home/20160111005488/en/Tensha-Therapeutics-
1053   Acquired-Roche (accessed on Apr 12, 2019).
1054   51. Perkmann, M.; Schildt, H. Open data partnerships between firms and universities: The
1055   role of boundary organizations. *Research Policy* **2015**, *44*, 1133–1143.
1056   52. Structural Genomics Consortium International Structural Genomics Consortium
1057   Announces $48.9 MILLION in Additional Funding to Continue the Search for New
1058   Medicines 2011.
1059   53. Montreal Neurological Institute and Hospital Open Science Available online:
1060   https://www.mcgill.ca/neuro/open-science-0 (accessed on Apr 12, 2019).
1061   54. Montreal Neurological Institute and Hospital Measuring The Impact Of Open Science
1062   Available   online:
1063   https://www.mcgill.ca/neuro/open-science-0/measuring-impact-open-science
1064   (accessed on Apr 12, 2019).
1065   55. SPOMAN Open Science About OS | Spoman OS.
1066   56. Weeber, M.; Vos, R.; Klein, H.; de Jong-van den Berg, L.T.W.; Aronson, A.R.;
1067   Molema, G. Generating Hypotheses by Discovering Implicit Associations in the
1068   Literature: A Case Report of a Search for New Potential Therapeutic Uses for
1069   Thalidomide. *J Am Med Inform Assoc* **2003**, *10*, 252–259.
1070   57. Houghton, J.; Sheehan, P. Estimating the Potential Impacts of Open Access to Research
1071   Findings. *Economic Analysis and Policy* **2009**, *39*, 127–142.
1072   58. Johnson, P.A.; Sieber, R.; Scassa, T.; Stephens, M.; Robinson, P. The Cost(s) of
1073   Geospatial Open Data. *Transactions in GIS* **2017**, *21*, 434–445.
1074   59. Huber, F.; Wainwright, T.; Rentocchini, F. Open data for open innovation: managing
1075   absorptive capacity in SMEs. *R&D Management* **2018**, *0*.
1076   60. Morgan, M.R.; Roberts, O.G.; Edwards, A.M. Ideation and implementation of an open
1077   science drug discovery business model – M4K Pharma. *Wellcome Open Research*
1078   **2018**, *3*, 154.
1079   61. Chataway, J.; Parks, S.; Smith, E. How Will Open Science Impact on
1080   University-Industry Collaboration? *Foresight and STI Governance (Foresight-Russia*
1081   *till No. 3/2015)* **2017**, *11*, 44–53.
1082   62. European Commission *Validation of the results of the public consultation on Science*
1083   *2.0: Science in Transition*; European Commission, 2015;
1084   63. Mueller-Langer, F.; Andreoli-Versbach, P. Open access to research data: Strategic
1085   delay and the ambiguous welfare effects of mandatory data disclosure 2017.
1086   64. Caulfield, T.; Harmon, S.H.; Joly, Y. Open science versus commercialization: a modern
1087   research conflict? *Genome Medicine* **2012**, *4*, 17.
1088   65. de Vrueh, R.L.A.; Crommelin, D.J.A. Reflections on the Future of Pharmaceutical
1089   Public-Private Partnerships: From Input to Impact. *Pharm Res* **2017**, *34*, 1985–1999.

66. Mark Ware Consulting Ltd *Access by UK small and medium-sized enterprises to professional and academic information*; Publishing Research Consortium, 2009;

67. Vines, T. Is There a Business Case for Open Data? Available online: https://scholarlykitchen.sspnet.org/2017/11/15/business-case-open-data/ (accessed on Apr 12, 2019).

68. Bilder, G. Crossref's DOI Event Tracker Pilot Available online: https://www.crossref.org/blog/crossrefs-doi-event-tracker-pilot/ (accessed on Apr 12, 2019).

69. Mowery, D.C.; Ziedonis, A.A. Markets versus spillovers in outflows of university research. *Research Policy* **2015**, *44*, 50–66.

70. Fukugawa, N. Knowledge spillover from university research before the national innovation system reform in Japan: localisation, mechanisms, and intermediaries. *Asian Journal of Technology Innovation* **2016**, *24*, 100–122.

71. Keseru, J. A new approach to measuring the impact of open data Available online: https://sunlightfoundation.com/2015/05/05/a-new-approach-to-measuring-the-impact-of-open-data/ (accessed on Apr 12, 2019).

72. Gold, E.R.; Ali-Khan, S.E.; Allen, L.; Ballell, L.; Barral-Netto, M.; Carr, D.; Chalaud, D.; Chaplin, S.; Clancy, M.S.; Clarke, P.; et al. An open toolkit for tracking open science partnership implementation and impact. *Gates Open Res* **2019**, *3*, 1442.

73. OpenMinted About. *OpenMinTeD* 2015.

74. Konfer About Konfer Available online: https://www.konfer.online/media (accessed on Apr 12, 2019).

75. Bishop Grosseteste University About LORIC – BGU.

76. Giovani, B. Open data for research and strategic monitoring in the pharmaceutical and biotech industry. *Data Science Journal* **2017**, *16*.

77. Houghton, J. *Open Access: What are the Economic Benefits? A Comparison of the United Kingdom, Netherlands and Denmark*; Social Science Research Network: Rochester, NY, 2009;