DEUTSCHE BUNDESBANK
EUROSYSTEM

On the value of data sharing
Empirical evidence from the Research Data and Service Centre

Technical Report 2023-08

This report was prepared for the 64th ISI World Statistics Congress
16 July 2023 - 20 July 2023 in Ottawa, Canada

The views expressed in this technical report are personal views of the authors and do not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem.

Deutsche Bundesbank, Research Data and Service Centre

Jannick Blaschke
Christian Hirsch

Research Data and Service Centre

Abstract

Access to timely and high-quality granular data is increasingly becoming a key factor for successful research projects. For accessing confidential administrative data, the introduction of Research Data Centres (RDCs) has been a big success story. RDCs are restricted-access facilities, often at the premises of the data owner, which provide accredited researchers with safe access to sensitive granular data. Although the benefits of an RDC are undisputed, there have been few attempts in the literature so far to measure them. Our impact assessment leverages on years of experience from one of the largest RDCs in Germany. We also discuss the challenges that we encounter during our impact assessment.

Keywords: rdc, rdsc, value, data sharing

Version: 1.0

Citation: Blaschke, J., and Hirsch, C. (2023). On the value of data sharing: Empirical evidence from the Research Data and Service Centre, Technical Report 2023-08 – Version 1.0. Deutsche Bundesbank, Research Data and Service Centre.

Contents

1 Introduction

Since the financial crisis more than a decade ago, central banks around the world have been increasingly collecting microdata on banks, firms, financial transactions and households. This treasure trove of data significantly supports their objectives for financial stability and monetary policy since interconnections between banks and the real economy are often so complex that many relationships only become visible in detailed, very granular data. At the same time, the availability of microdata holds an enormous potential for the academic research community outside central banks. The European System of Central Banks (ESCB) even explicitly enshrines this additional use case in its legal basis for the collection of statistical data[1].

The basic prerequisite for using confidential microdata for external research projects is that confidentiality and, in particular, the protection of reporting agents, is maintained. For this purpose, a mixture of anonymisation techniques and organisational measures is usually taken. The "five safes" framework very aptly describes these measures (Desai et al. 2016; Ritchie, 2017). It distinguishes safe projects (Is this use of the data appropriate?), safe people (Can the researchers be trusted to use it in an appropriate manner?), safe data (Is there a disclosure risk in the data itself?), safe settings (Does the access facility limit unauthorised use?), and safe outputs (Are the statistical results non-disclosive?).

For the practical implementation of the "five safes", there are different approaches that vary depending on the confidentiality level and the technical and staff equipment of the data provider. The challenge each time is to find the right balance between confidentiality and data usability. The literature[2] usually distinguishes four access modes:

1. **Secure on-site access:** Data provision at the premises of the institution in a dedicated secure environment where researchers cannot use the internet or their own laptops, tablets or phones.
2. **Scientific Use Files (SUFs):** Off-site provision of data, usually via download. In contrast to Public Use Files, SUFs are not absolutely (fully) anonymised.
3. **Remote execution:** Researchers send code and the data-providing institution sends back results. The researchers themselves do not have access to confidential microdata.
4. **Remote access:** Researchers can access data remotely via a secure IT solution from their own institution.

For accessing confidential administrative data, the introduction of Research Data Centres (RDCs) has been a big success story. RDCs are restricted-access facilities which steer and provide access to confidential microdata for independent research. However, since the operation of an RDC is also associated with financial and personnel costs for the institution providing the data, the question of its net value (benefits minus costs) inevitably arises.

We draw from the literature on the evaluation of research outcomes (see Banzi et al, 2011; and Greenhalgh et al., 2016, for reviews of this literature) when calculating the value of running an RDC. This literature stems from the interest of funders and policy makers to assess the benefits of

1 [The ESCB shall use confidential statistical information exclusively for the exercise of the tasks of the ESCB except in any of the following circumstances:] (…) (c) for granting scientific research bodies access to confidential statistical information which does not allow direct identification, and with the prior explicit consent of the authority which provided the information;" Council Regulation (EC) No 2533/98 of 23 November 1998 concerning the collection of statistical information by the European Central Bank, Article 8(1c).
2 E.g. INEXDA (2020); Desai, et al. (2016) or Nordholt (2021).

funding research projects (e.g. Lane, 2009; Banzi et al., 2011), which contributes to the accountability of researchers and informs future allocation decisions (Greenhalgh et al., 2016) (Greenhalgh, Raftery, Hanney, and Glover (2016)). Ideally, we would like to operationalise the analytical approach in Bender, Blaschke, and Hirsch (2023) Bender et al. (2023), where value is calculated as (direct and indirect) benefits minus (direct and indirect) costs.

They argue that the value of an RDC is essentially generated through three channels: (1) Added value for researchers (e.g. enabling a Ph.D. thesis), (2) added value for the public (e.g. knowledge gain through published research), and (3) added value for the data-providing institution (e.g. improved data quality through feedback loops from the researcher to the data producing business unit). Measuring the direct costs associated with running an RDC, such as staff costs or IT equipment costs, is relatively straightforward[3)]. In addition, any expected costs occurring in the unlikely event of a major data leak due to malicious behaviour of a researcher can be categorised in risk classes that roughly predict the expected costs[4)].

This approach shares similarities with monetarisation frameworks (e.g. Johnston et al., 2006) used, for example, to evaluate investments in clinical trials on public health. There are three main drawbacks of this approach related to assumptions, causality, and the field of study. First, the result of this approach hinges largely on the assumptions put into the analysis. For example, in the case of RDCs, one important assumption concerns the choice of measurement of the added value to researchers. This could, for example, be measured by outcomes such as the researcher earning a degree or obtaining a publication in a peer-reviewed journal[5)].

Both can be measured by a simple dummy variable that is either yes (e.g. researcher obtained a degree) or no (e.g. she did not obtain a degree). However, the value of such an outcome is not easily quantifiable. Freier et al. (2015) find that a master's degree leads to 14% higher earnings for law students while Swindler and Goldreyer (2002) estimate the lifetime present value of a publication to be between about 20,000 and 34,000 US dollars. Results of the study would be much different when using the lower or upper bound or assuming different levels from which to add the 14% increase.

Second, even for seemingly well objectifiable and quantifiable outcomes like publications, it is not straightforward to determine the exact contribution the provision of confidential microdata had, e.g. to a publication in a peer-reviewed journal. Had the publication not been possible without the confidential microdata? Had the publication been possible (perhaps in a modified form) with a dataset from a commercial data vendor? Thus, one could explore whether there are any substitutes to the provided microdata and if so, how well the outcome could be achieved using these substitute data. Furthermore, substitutes may be available but at a cost, while using data from the RDC is free of charge.[6)]

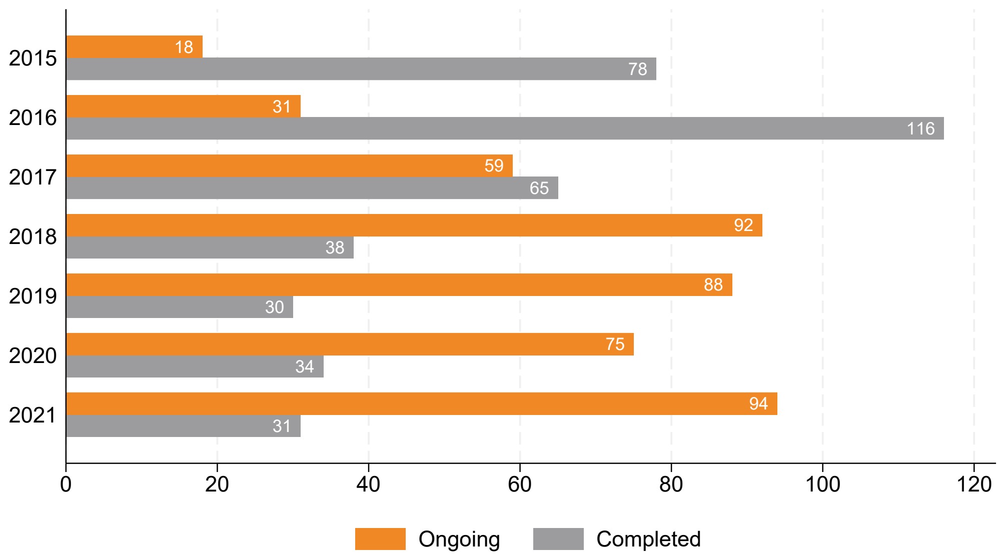
Third, the field of study of this paper (economics) makes attribution of value complex, as it lacks

3 Examples include salary of FTEs working at the RDC, PCs provided to the researchers, costs for office space, etc.
4 Such an evaluation of the expected costs and their likelihood has usually already been done when assessing the security of IT systems in which such data are stored, processed or analysed.
5 In addition, one can use registered patents as a proxy for general innovation. One example of this is Chevalier et al. (2020). However, since in this paper we mainly refer to financial and economic research, patents do not play an important role in our approach.
6 The RDC does not charge researchers for using its data. If data is secure, on-site access researchers may incur costs for traveling to the secure site, for example.

"the bottom-up" (Greenhalgh et al, 2016) quality of health research. In health research, costs of clinical trials can be added up and subtracted from the value of a new treatment. The issue is that the relation between economic research and the non-academic sphere is non-linear (Lane, 2009) in nature due to complex spillover effects, which are hard to capture in the data.

We therefore take a different approach and utilise the payback framework (Buxton and Hanney, 1996; Rollins et al, 2020) to evaluate the benefits of RDCs. The payback framework proposes to evaluate research outcomes along five dimensions which aim to capture direct as well as indirect outcomes (or paybacks) of research. We choose this framework because it is commonly used in the literature (Banzi et al, 2011; Greenhalgh et al., 2016). Furthermore, it goes well together with our database that we explain in detail in Section 2, although there are some challenges that we explain in Section 3.

The aim of this paper is to make a first attempt to evaluate the outcomes of research projects from an RDC, leveraging from the experiences and actual data on past and current research projects from the Deutsche Bundesbank's Research Data and Service Centre (RDSC)[7]. Section 2 presents the data used for this paper and shows the work of the RDSC through a series of metrics. Section 3 calculates the impact of data sharing on researchers. Finally, Section 4 summarises the results and gives an outlook on possible future analyses.



Source: Research Data and Service Centre of Bundesbank, as at December 2022

Figure 1: Number of research projects by status

7 We encourage all readers interested in a more detailed description of the RDSC to refer to our technical report "Data Access to Micro Data of the Deutsche Bundesbank" (Schönberg, 2019). In addition, information on the RDSC can be found on our website https://www.bundesbank.de/en/bundesbank/research/rdsc.

2 Data and descriptive statistics

We base our results on data from the RDSC's research project database, where our team maintains metadata information for each research project, such as the project description, requested datasets, participating researchers, status of the project and all publications resulting from the project. The collection of this information is a core part of the daily work at the RDSC and is regularly reviewed and updated.

For the key figures and graphs shown in the remainder of this paper, we only consider projects in the period from January 2015 to December 2021. This is because 2015 is the first full year of the RDSC's launch in mid-2014. We chose to exclude projects that started after December 2021 because we are interested in the outcome of research projects and we want to let projects have sufficient time to produce an outcome (Rollins et al., 2020). This leaves us with 849 projects.

Note that we only include projects that have all information necessary for our analysis, have been through an application process and have since been closely monitored by Bundesbank staff. This requirement excludes from our analysis the large number of data available on the Bundesbank website[8)]. These datasets comprise aggregated data and are accessible without filing an application with the RDSC. Therefore, we do not have information on the outcomes and the researchers using this particular type of data.

Figure 1 presents the number of ongoing and completed projects broken down by the year in which the project started in our database. In this context, "Completed" means that the project members have explicitly declared that they have finished the project. Note that completed does not necessarily mean that there is a publication. We will come back to this in the next section. Please also note that all results are valid as at December 2022 when we started this analysis. Towards the end of our sample period the number of ongoing projects increases while the number of completed projects decreases, as one would expect. The overall number of projects stays roughly constant over the years, with the exception of 2015 where the number of projects was smaller compared to the other years.

The RDSC provides standardised access to around 30 selected micro datasets[9)] collected by the Bundesbank in accordance with its statutory mandate to be used in independent scientific research projects. However, this is not a stable number as new datasets are being collected by the ESCB and are in turn made available to researchers as time goes by. The RDSC provides microdata predominantly via secure on-site access or SUFs, i.e. offsite[10)]. Which access mode is available in each case depends on the data selection, since, for example, only certain datasets can be shared as SUFs.

Figure 2 presents the distribution of projects by year and access mode. In all years, the number

8 See https://www.bundesbank.de/en/statistics/time-series-databases

9 For an overview to all datasets that external researchers can request, please see the list on our website: https://www.bundesbank.de/en/bundesbank/research/rdsc/research-data. Please note that we provide our internal staff with additional datasets that, due to legal restrictions, externals are not entitled to request.

10 Due to the increased difficulty for researchers of remote execution, the RDSC affords this option only to those researchers who have demonstrated sufficient experience of the respective microdata in a given project. Thus, all remote execution projects started as secure on-site projects. For this reason and due to the generally small number of remote execution projects (21 active projects as of December 2022), we distinguish in Figure 2 only between secure on-site access and SUFs.

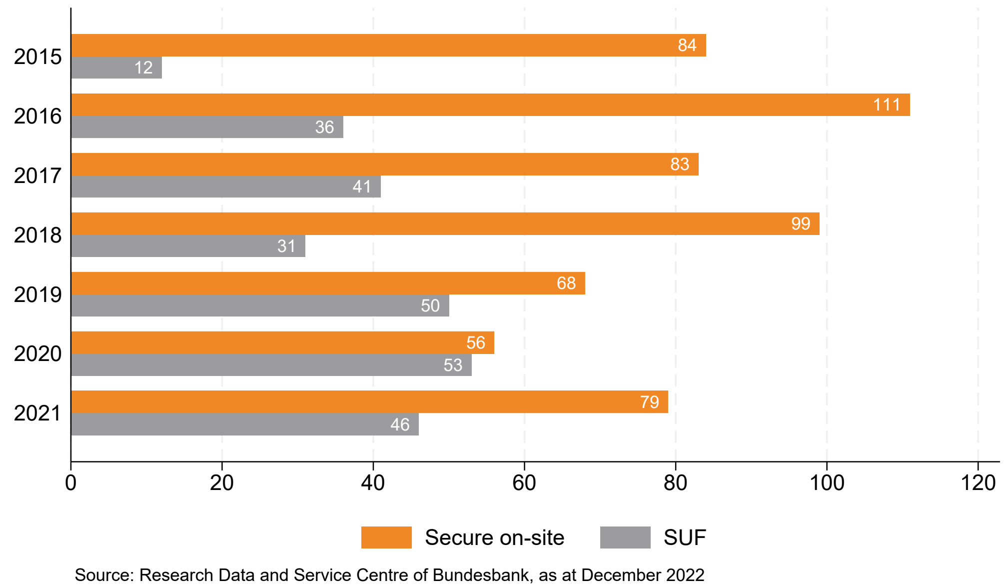


Figure 2: Number of research projects by access mode

of secure on-site access projects exceeds the number of SUF projects. The magnitude of this difference varies and begins to shrink as of 2019. There are at least two reasons for this. The first reason is the Covid-19 pandemic, which severely limited travelling, especially in 2020 but also at the start of 2021. Consequently, we observe a shift of researchers towards SUFs, although the number of projects in the database does not decline by much compared to pre-pandemic years. Second, the RDSC made available a new dataset as an SUF[11)] during the year 2019, which also contributed to an increase in the number of SUF projects in the database.

Since the RDSC facilitates data access for both Bundesbank-internal and external researchers, it is also worth looking at the distribution of the researcher type across projects. Of the 849 projects, 162 (19.1%) are purely internal, 549 (64.7%) are purely external and 138 (16.2%) are mixed projects where Bundesbank staff members collaborate with external researchers. Furthermore, we observe that projects are conducted by one researcher in about 50% of projects in our database. The mean number of researchers per project is roughly 2.

Next, we turn to analysing the research area of projects in our database. We construct a variable for data type that clusters datasets into broader categories based on their contents. For example, the Monthly Balance Sheet Statistics (BISTA) (Gomolka et al, 2022) falls into the data type category banks because it provides information on balance sheets of banks.[12)] Figure 3 shows projects by researcher type and data type. The RDSC provides microdata on banks, companies, securities and households for researchers. Company and household data are most popular with external researchers, while projects with internal participation tend to use securities and bank data. Note that projects could use data from more than one data type.

11 Bundesbank Online Panel – Households (BOP-HH). For more information see https://www.bundesbank.de/en/bundesbank/research/rdsc/research-data/bop-hh-757542
12 For examples of allocating datasets to data types, see https://www.bundesbank.de/en/bundesbank/research/rdsc/research-data

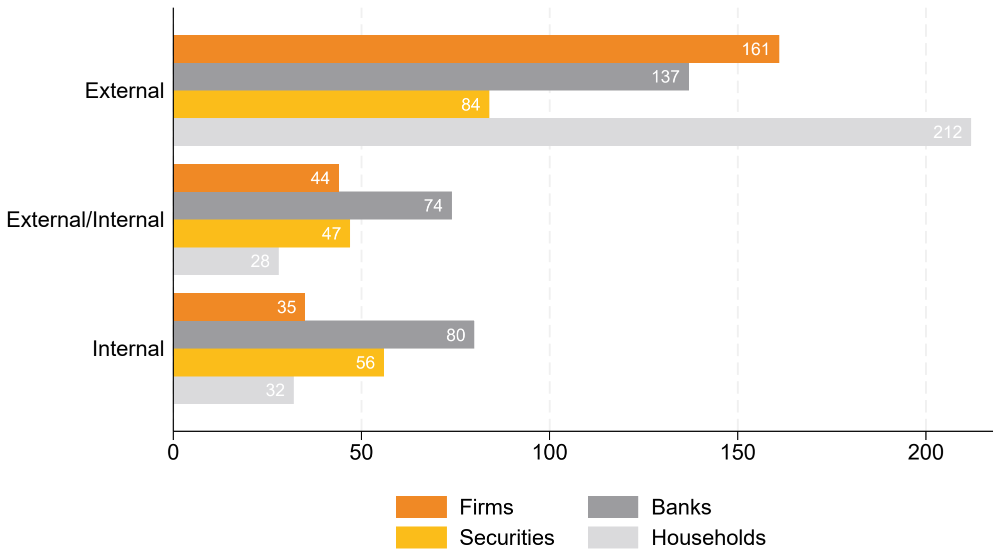


Figure 3: Number of research projects by researcher type and data type

Turning to the number of datasets, about 60% of projects use one dataset for their project. This number goes down to 40% if we exclude SUFs, which cannot be merged[13)] with other datasets and are therefore mostly used individually. Furthermore, 28% of all projects request more than three different datasets if we exclude SUFs[14)]. The average number of non-SUF-datasets requested for on-site access per project is 2.09. The average number goes up to 2.70 if we only include projects where researchers have actually started working with the data. This difference shows that assumptions about the status of the project play an important role when providing summary statistics. This will become important in the next section.

In summary, the RDSC has facilitated many new research projects. While internal projects were also possible before its establishment[15)], albeit through a less streamlined process, it is mainly the 549 purely external and the 138 mixed projects that account for the value added. It is also noticeable that researchers not only choose the comparatively more convenient access mode via SUFs, but also travel physically to the four RDSC sites in order to use our microdata. This is particularly remarkable, as many researchers travel from abroad to access our microdata.

13 SUF data do not include any identifier that could be used to link them to other data.
14 Note that we exclude SUFs in this calculation because SUF projects usually use only one dataset.
15 Since the establishment of the RDSC, internal researchers have benefited in particular from the provision of standardised research datasets, detailed documentation as well as record linkage methods (Schild et al, 2017).

3 Evaluating the impact of data sharing

We base our evaluation on the benefits of RDCs on the payback framework (Buxton and Hanney, 1996; Rollins et al., 2020). We follow Sorensen et al. (2021) and organise the framework's five dimensions along a timeline that starts from the analytical work of the researcher. Closest in time to this research analysis are "immediate outcomes" such as publications, seminar presentations or earning a degree (PhD or master's). Next in line are "intermediate outcomes" comprising the dissemination to and use of research in policy and practice. Hence, "immediate outcomes" are confined to the academic sphere where "intermediate outcome" attempt to measure value generated from cross-sector knowledge transfer (Sorensen et al., 2021). Finally, furthest from the research analysis are "end outcomes" (e.g. societal impact), which are confined to the non-academic sphere.

Deciding what to measure and how to measure it gets more challenging the further we move away from the research analysis. Indeed, the literature seems to disagree on both the definition and measurement of societal impact (i.e. "end outcomes") (Lane, 2009; Bornmann, 2013; Sorensen et al.; 2021, Bührer et al., 2022). On the other end of the timeline, there seems to be a consensus to use some form of bibliometric information to measure the impact of "immediate outcomes" (Banzi et al., 2011; Wilsdon et al., 2017; Rollins et al., 2020).[16] The "intermediate outcomes" fall somewhere in the middle, although Sorensen et al. (2021) argue that they tend to lean towards "end outcomes" as there are no generally accepted measurements available in the literature.

This challenge leads to a situation in which the literature uses quantitative measures when assessing "immediate outcomes" but uses qualitative and descriptive measures when assessing "intermediate" and "end outcomes" (Rollins et al., 2020; Sorensen et al., 2021). Part of this can be explained by time. Users may want to assess research projects in close proximity to the analytical work of the researchers but end outcomes may take time to fully materialise.[17] (Lane, 2009; Rollins et al., 2020).

The field of study in this paper further intensifies this challenge "because the creation and transmission of knowledge and technologies result from complex human and social interactions" (Lane, 2009, p.1274). Indeed, research in economics rarely leads to linear (Lane, 2009) outcomes such as patents or products (Banzi et al., 2011; Rollins et al., 2020) or other easily measureable "intermediate outcomes" regularly used in health care assessments.

Finally, the literature deals with this challenge by either limiting their assessment to "immediate outcomes" in the academic sphere (e.g. Weinberg et al., 2014; Chevalier et al., 2020) or by reporting answers from researcher surveys to gauge, for example, how the research affects workplace use by managers or other employees (Sorensen et al., 2021). As we draw our data from the project database of the Bundesbank's RDSC, we do not have access to results from such surveys, although we find that users change sectors once they have concluded their research project, which is indicative of some sort of knowledge transfer (Sorensen et al., 2021).

In summary, our data is well suited for assessing "immediate outcomes" that materialise in the academic sphere, but imperfect when assessing outcomes that attempt to measure the trans-

16 Weinberg et al. (2014) uses employment and payroll records to measure short-term effects.
17 In our case the shortest elapsed time between the start of a project and the assessment is one year. We conduct our assessment at the end of 2022 and consider projects that start in 2021.

fer of knowledge from the academic to the non-academic sphere and the value creation in the non-academic sphere. We argue above that we share this challenge with the impact literature. However, part of this challenge also arises from applying this literature to economics and the data that we use for this study.

We now turn to presenting the results of our assessment. The payback framework differentiates between two types of "immediate outcomes": knowledge production as well as research targeting and capacity building (Banzi et al., 2011; Rollins et al., 2020). We start with the former, which we measure by the number of projects that produced a publication. All researchers are obliged to submit their publication to us for acceptance before publication (RDSC, 2021). Those who are working with SUFs have to submit at least a specimen copy of their publication. Therefore, our database is well suited for calculating the number of publications stemming from data sharing. We use the following methods to define publications.

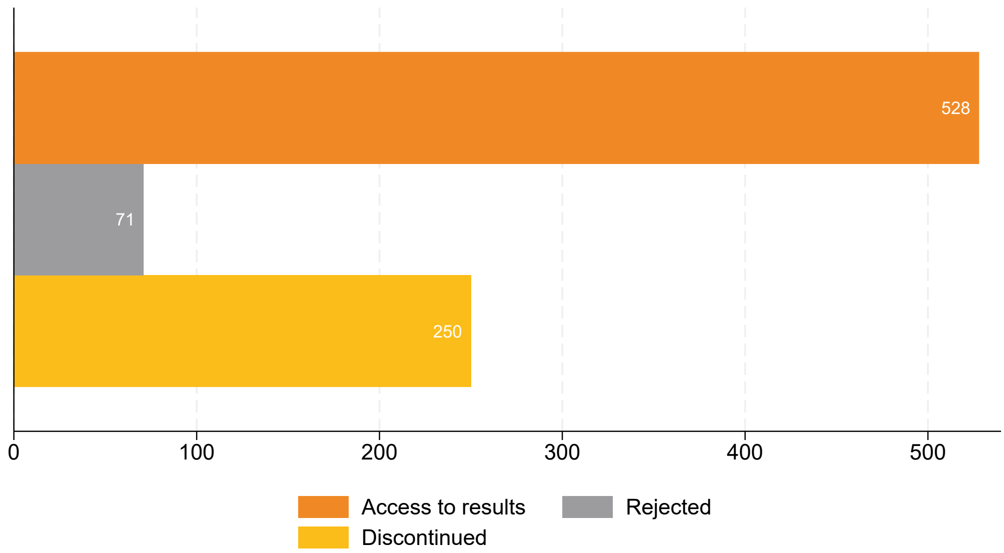
First, we follow the extant literature (Rollins et al., 2020) and consider papers published in peer-reviewed academic journals as a publication. Second, we further define working papers as publications. This choice is motivated by the fact that Banzi et al. (2011) count conference presentations as knowledge production and researchers usually present their working papers at conferences before submitting to a journal. Furthermore, it also takes into account the relatively short time period between research analysis and our assessment (remember our sample goes until December 2021), which may hinder papers getting published because the process of publishing in a peer-reviewed journal takes some time.

Publications can only result from projects for which at least one output was released by the RDSC or for which the data was provided as an SUF. Projects that the RDSC rejected in the application phase[18)], or where the researchers stopped the process before their first visit to the RDSC (or before the SUF was sent out)[19)], cannot have a publication. Figure 4 provides a breakdown of all projects according to whether the researchers in the project accessed data. Out of the 849 projects in our database between 2015 and 2021, 528 are eligible to produce outcomes. In contrast, 321 projects will not produce any publications as either they are discontinued or the application was rejected. If we exclude all projects for which no output was possible, we see in Figure 5 that of 528 projects, 164 (31.1%) led to a publication and 364 (69.9%) did not.

When taking a closer look at the 364 projects that had access to results but have not published anything yet, we find that 147 (40.38%) are projects where the lead researcher is a graduate student (i.e. master's or PhD student)[20)]. This makes sense, as graduate students are less likely to pursue publications in peer-reviewed journals.

The second immediate outcome in the payback framework refers to research targeting and capacity building. Banzi et al. (2011) report in their literature review article that the main proposed

18 Possible reasons for rejection of a research project are conflicts of interests. We do not reject projects due to content-related reasons.
19 We do not have a statistic of why researchers do not follow through with their projects. Anecdotal evidence suggests that reasons include the realisation during the application process that data are insufficient to answer the research question or a shift towards other projects. Of course, travel restrictions during the pandemic also contributed to projects being discontinued before the first research visit.
20 We determine whether the project is a Bachelor, Master or PhD thesis based on the status of the lead researcher at the time of the project start. The lead researcher is our contact person in the project and usually also the person who submits the first project proposal.



Source: Research Data and Service Centre of Bundesbank, as at December 2022

Figure 4: Number of research projects by project status

indicators are related to "new research lines or know how; career promotions, PhD, masters" (Banzi et al., 2011, p 3). Rollins et al. (2020) utilise answers from a survey to assess research lines and know how. Since we have no means of conducting a survey, we limit our assessment to degrees (PhD and master's).

We use the following method to gather this information. First, we identify master's or PhD students from their application. In order to gain access to data, researchers need to complete an application form where they need to report their current position among other information. For the purpose of this assessment, we only select applications where the position of the lead researcher is either a master's or PhD student. Note that this will likely lead to a lower bound estimate insofar as researchers other than the lead researcher could very well also be graduate students.

Second, we search for the name of the researcher and manually assess whether they have completed their degree. One drawback of this method is that the likelihood of finding a name depends on how common the name is. Note that we only do this for completed research projects because we assume that the researcher needs the paper from the project as part of his thesis. This method reveals that the 528 projects with access to results produced 127 completed degrees.

Measuring "intermediate" or "end outcomes" is, as we explain above, not feasible for this paper for reasons inherent to the literature, the field of study and the data that we use. However, the manual search described above provides us with a rough proxy for "intermediate outcomes", which measure the transfer of knowledge from the academic to the non-academic sphere.

Sorensen et al. (2021) use movement between academia and industry (i.e. non-academia) to capture knowledge exchange. When assessing whether a researcher has completed a degree, we also record whether his new employee is an academic institution or not. We find that 76 lead researchers move to industry, supporting the notion that the data sharing in a RDC also leads to

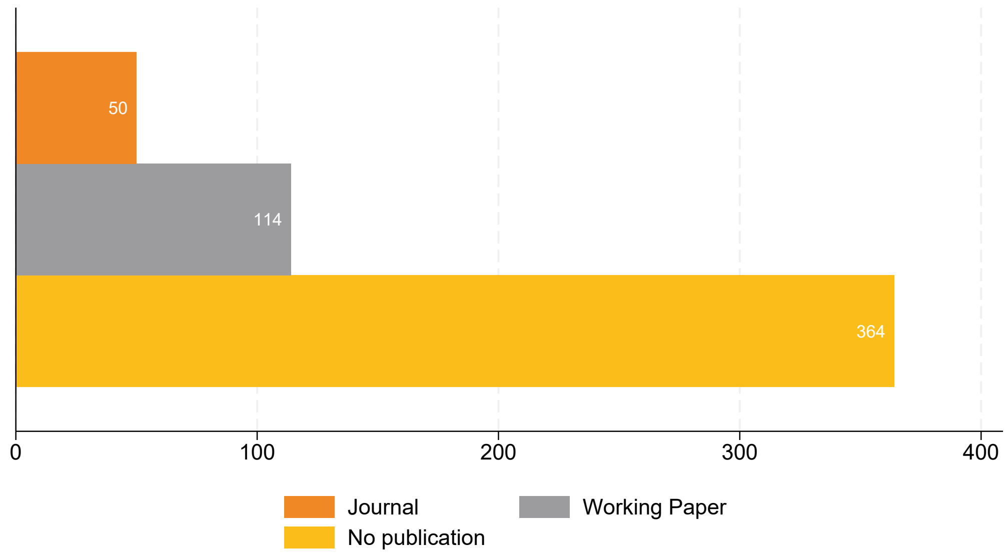


Figure 5: Number of publications

“intermediate outcomes”.

4 Conclusion

In this paper, we have made a first attempt to assess the value of an RDC, examining data from the research project database of the Deutsche Bundesbank's RDSC. For this, we take the payback framework commonly used to assess health care research projects and adapt it to the field of economics. We find that data sharing leads to immediate outcomes such as publications and completed degrees.

We do not attempt to quantify other benefits of data sharing such as its societal value, which may exceed the direct benefits to researchers as it potentially affects more people. Therefore, readers should interpret our results here as a lower bound. In addition, we do not attempt to measure the contribution of data to publications.

Our assessment also does not take into account the fact that some of the datasets can also be accessed from commercial vendors for a fee. By contrast, data sharing at the Deutsche Bundesbank's RDSC is free of charge. It would be interesting to analyse whether this contributes to levelling the playing field among students in Germany and beyond.

References

Banzi, R., Moja, L., Pistotti, V., Facchini, A., and Liberati, A. (2011). Conceptual frameworks and empirical approaches used to assess the impact of health research: An overview of reviews. *Health Research Policy and Systems*, *9*(26).

Bender, S., Blaschke, J., and Hirsch, C. (2023). Statistical data production in a digitized age: The need to establish successful workflows for micro data access. *Advances in Business Statistics, Methods and Data Collection*, 519–536.

Bornmann, L. (2013). Concepts and methods to measure societal impacts – an overview. *Advances in Information Science*, *64*(2), 217–233.

Buehrer, S., Feidenheimer, A., Walz, R., Lindner, R., Beckert, B., and Wallwaey, E. (2022). Concepts and methods to measure societal impacts– an overview. *Fraunhofer ISI Discussion Papers Innovation Systems and Policy Analysis*, *74*.

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

Chevalier, G., Chomienne, C., Jeanrenaud, N. G., Lane, J., and Ross, M. (2020). A new approach for estimating research impact: An application to french cancer research. *Quantitative Science Studies*, *1*(4), 1586–1600.

Desai, T., Ritchie, F., and Welpton, R. (2016). Five safes: Designing data access for research. *Economics Working Paper Series No 1601*.

Freier, R., Schumann, M., and Siedler, T. (2015). The earnings returns to graduating with honors—evidence from law graduates. *Labour Economics*, *Volume 34*, 39–50.

Gomolka, M., Schäfer, M., and Stahl, H. (2022). *Monthly Balance Sheet Statistics (BISTA), Data Report 2022-08 – Metadata Version BISTA-Doc-v4-0, Data Report 2022-08*. Deutsche Bundesbank, Research Data; Service Centre. Retrieved from https://www.bundesbank.de/resource/blob/745176/8473d9971799d58295cbe340a8199e06/mL/2022-08-bista-data.pdf

Greenhalgh, T., Raftery, J., Hanney, S., and Glover, M. (2016). Research impact: A narrative review. *BMC Medicine*, *14*(1), 1–16.

INEXDA. (2020). *Final Report of the INEXDA Working Group on Data Access, Version 1.0*. International Network for Exchanging Experience on Statistical Handling of Granular Data. Retrieved from https://www.inexda.org/wp-content/uploads/2020/05/FINAL_INEXDA_WG_DA_Final_Report.pdf

Johnston, S. C., Rootenberg, J. D., Katrak, S., Smith, W. S., and Elkins, J. S. (2006). Effect of a US national institutes of health programme of clinical trials on public health and costs. *The Lancet*, *367*(9519), 1319–1327.

Lane, J. (2009). Assessing the impact of science funding. *Science*, *324*(5932), 1273–1275.

Nordholt, E. S. (2021). *Access to different kinds of statistics netherlands' microdata.* UNECE Conference of European Statisticians - Expert Meeting on Statistical Data Confidentiality 1-3 December 2021, Poland. Retrieved from https://unece.org/sites/default/files/2021-12/SDC2021_Day1_Schulte%20Nordholt_AD.pdf

Research Data and Service Centre. (2021). *Rules for visiting researchers at the RDSC, Technical Report 2021-02 - Version 1-0*. Deutsche Bundesbank, Research Data; Service Centre. Retrieved from https://www.bundesbank.de/resource/blob/826176/ffc6337a19ea27359b06f2a8abe0ca7d/mL/2021-02-gastforschung-data.pdf

Ritchie, F. (2017). The "five safes": A framework for planning, designing and evaluating data access solutions. *Data for Policy 2017: Government by Algorithm? (Data for Policy). Zenodo*.

Rollins, L., Llewellyn, N., Ngaiza, M., Nehl, E., Carter, D. R., and Sands, J. M. (2021). Using the payback framework to evaluate the outcomes of pilot projects supported by the georgia clinical and translational science alliance. *Journal of Clinical and Translational Science*, *5*(1), e48.

Schild, C. J., Schultz, S., and Wieser, F. (2017). *Linking Deutsche Bundesbank Company Data using Machine-Learning-Based Classification. , Technical Report 2017-01*. Deutsche Bundesbank, Research Data; Service Centre. Retrieved from https://www.bundesbank.de/resource/blob/604828/f3227be248777d3f414b89b395e875cd/mL/2017-01-company-data.pdf

Schönberg, T. (2019). *Data Access to Micro Data of the Deutsche Bundesbank, Technical Report 2019-02*. Deutsche Bundesbank, Research Data; Service Centre. Retrieved from https://www.bundesbank.de/resource/blob/801044/6484d9e4aa2be3610b7378a48a1916de/mL/2019-02-data-access-data.pdf

Sørensen, O. H., Bjørner, J., Holtermann, A., Dyreborg, J., Sørli, J. B., Kristiansen, J., and Nielsen, S. B. (2022). Measuring societal impact of research—developing and validating an impact instrument for occupational health and safety. *Research Evaluation*, *31*(1), 118–131.

Swidler, S., and Goldreyer, E. (1998). The value of a finance journal publication. *The Journal of Finance*, *53*(1), 351–363.

Weinberg, B. A., Owen-Smith, J., Rosen, R. F., Schwarz, L., Allen, B. M., Weiss, R. E., and Lane, J. (2014). Science funding and short-term economic activity. *Science*, *344*(6179), 41–43.

Wilsdon, J., Bar-Ilan, J., Frodeman, R., Lex, E., Peters, I., and Wouters, P. (2017). *Next-generation metrics: Responsible metrics and evaluation for open science*.