Research Policy xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Open access to data: An ideal professed but not practised

Patrick Andreoli-Versbach[a,b,*], Frank Mueller-Langer[a,b]

[a] Max Planck Institute for Innovation and Competition, Marstallplatz 1, D-80539 Munich, Germany
[b] International Max Planck Research School for Competition and Innovation (IMPRS-CI), Marstallplatz 1, D-80539 Munich, Germany

ARTICLE INFO

Article history:
Received 4 April 2013
Received in revised form 11 April 2014
Accepted 15 April 2014
Available online xxx

JEL classification:
B40
C80
L59

Keywords:
Voluntary data-sharing
Mandatory data-disclosure
Open science

ABSTRACT

Data-sharing is an essential tool for replication, validation and extension of empirical results. Using a hand-collected data set describing the data-sharing behaviour of 488 randomly selected empirical researchers, we provide evidence that most researchers in economics and management do not share their data voluntarily. We derive testable hypotheses based on the theoretical literature on information-sharing and relate data-sharing to observable characteristics of researchers. We find empirical support for the hypotheses that voluntary data-sharing significantly increases with (a) academic tenure, (b) the quality of researchers, (c) the share of published articles subject to a mandatory data-disclosure policy of journals, and (d) personal attitudes towards "open science" principles. On the basis of our empirical evidence, we discuss a set of policy recommendations.

1. Introduction

Sharing and granting access to data (disclosure) enables researchers to replicate, verify and expand existing works. In addition, it helps to deter and detect data-related fraud and creates incentives for compiling high-quality data and codes, thereby reducing errors.[1] These arguments would seem to make voluntary data-sharing "a requirement for responsible scholarship" (Journal of Political Economy, 1975, p. 1296). In addition, most researchers seem to embrace the idea of self-correction and replicability in science – which is immediately associated with data-sharing (Nelson, 2009).

There is ample anecdotal evidence, however, that the ideal is not being followed in practice due to disincentives to share data (Moffitt, 2007). Voluntary data-sharing is also not part of the norms in many fields in which non-disclosure of data is to some extent an accepted practice (Tenopir et al., 2011). Furthermore, researchers have personal incentives to avoid sharing of intermediate research results (Thursby et al., 2009) and the willingness to share depends on whether the information is shared privately in response to a request or publicly, e.g., in academic presentations or on the Internet (Haeussler et al., 2014). In addition, according to Anderson et al. (2008, p. 99), the "'supply of' replicable results in economics has been minimal". As Hamermesh (2006, p. 715) claims, "economists treat replication . . . as an ideal to be professed but not to be practised". However, there is very little systematic evidence on the status quo and the key drivers of data-sharing that goes beyond anecdotes or intuitive insights – both of which are not particularly suitable foundations for editorial or public policies.

The aim of this paper is twofold. First, we provide systematic empirical evidence that data-sharing and facilitated access to data and codes in economics and management is scarce. Although the ease of sharing has increased recently and although many top-ranked journals in economics and management nowadays require data-sharing, researchers tend to publish neither data nor codes on their websites or in public data repositories. We constructed a unique data set describing the data-sharing behaviour of 488 randomly selected empirical researchers affiliated with the top 100 economics departments and top 50 business schools. In our sample, 82 empirical researchers (16.8%) sporadically share data, while only 12 (2.46%) share data in a comprehensive and clear way. The large majority, 394 (80.74%), neither share data nor provide any indication of whether or where the data is available.

* Corresponding author at: Max Planck Institute for Innovation and Competition, Marstallplatz 1, D-80539 Munich, Germany. Tel.: +44 7586312838.
E-mail addresses: patrick.andreoliversbach@gmail.com, patrick.andreoli-versbach@imprs-ci.ip.mpg.de (P. Andreoli-Versbach), frank.mueller-langer@ip.mpg.de (F. Mueller-Langer).

[1] The benefits of releasing data are not confined to replication of results or new research. For instance, they may also help to reduce corruption (Ferraz and Finan, 2008) or detect collusion (Christie and Schultz, 1994).

http://dx.doi.org/10.1016/j.respol.2014.04.008
0048-7333/

Second, based on the theoretical literature we derive testable hypotheses on data-sharing in economics and management. Our data allow us to relate voluntary data-sharing via researchers' websites and public data repositories to observable characteristics of the respective researchers and their institutions.[2] Our regression analysis allows us to show that voluntary data-sharing is systematically associated with three factors: personal incentives, institutional factors and personal attitudes towards "open science". Our empirical results confirm these theoretical predictions and show that the likelihood of voluntary data-sharing increases with (a) academic tenure, (b) the quality of researchers, (c) the share of published articles that are subject to a mandatory data-disclosure policy of journals and (d) personal attitudes towards open science. Based on our results, we derive a set of implications for journal and university policy to address the current low sharing status and encourage data-sharing.

The debate on data-sharing has recently gained momentum due to the data fraud scandal concerning the social psychologist Diederik Stapel from Tilburg University. Considered a star scientist in the Netherlands and abroad, his studies on social behaviour attracted wide press coverage. In 2012, the Levelt Committee et al. (2012, p. 17) found that he had committed scientific fraud including "fabrication, falsification or unjustified replenishment of data" in 55 publications. They concluded that the data that underlie publications "must remain archived and be made available on request to other scientific practitioners" in order to increase detection and to discourage fraud (Levelt Committee et al., 2012, p. 58). This debate has been further pursued in Nature (2009, p. 145), whose editors claim that "research cannot flourish if data are not preserved and made accessible" and "data management should be woven into every course in science, as one of the foundations of knowledge".

Besides fraud, the debate surrounding the coding error by two leading economists on the negative relation between debt and growth, which has been used to justify austerity programmes around the world, has also advanced the debate on data-sharing and replication (Herndon et al., 2013; Reinhart and Rogoff, 2013).[3]

This paper relates to two intertwined strands of literature which focus on the personal and institutional factors of data-sharing and replication in applied economics (Dewald et al., 1986; Feigenbaum and Levy, 1993; McCullough et al., 2006) and on the incentive structure for sharing of intermediate research results (Dasgupta and David, 1994; Haeussler, 2011; Haeussler et al., 2014).

The first strand of literature emphasizes that inadvertent errors do occur in empirical publications (Anderson et al., 2008; Dewald et al., 1986; McCullough et al., 2006). However, in applied economics, there is not a strong tradition of data-sharing or of replication studies for which the availability of data is a prerequisite (McCullough et al., 2006).[4] As Anderson et al. (2008, p. 101) put it, "the incentive structure of publish-or-perish is just too strongly skewed toward irreproducibility". Further, as the publication market for replication studies is limited, investing time and effort in writing a replication study is not an efficient use of a researcher's resources (Hamermesh, 1997; McCullough et al., 2006; Mirowski and Sklivas, 1991). Consequently, as replication of a researcher's results is highly unlikely to occur, it is also not rational to invest significant time and effort in ensuring the replicability of research results, e.g., by compiling data in a reproducible way (McCullough et al., 2006). Feigenbaum and Levy (1993) suggest that rational researchers are reluctant to share data in order to delay or even prevent attempts to replicate their results. Keeping research data secret may also be a rational strategy for the creators of the data to protect their competitive advantage as sharing of data would lower the competing researchers' cost of recollecting the data and rewriting the code (McCullough, 2009). In particular, the original creator of research data may have an incentive to keep the data secret until their private value is fully exploited in subsequent publications (Anderson et al., 2008; Stephan, 1996). In addition, Anderson et al. (2008, p. 99) argue that the" 'respect for the scientific method' is not sufficient to motivate . . . editors of professional journals to ensure the replicability of published results". While some of the top-ranked economics journals have recently introduced data availability policies to address this issue, the vast majority of journals either do not have a policy that requires authors to share their data or are reluctant to enforce it (McCullough, 2009; McCullough and Vinod, 2003). Also, Anderson et al. (2008) put forward that authors may generally be hesitant to share their data and code despite their pre-publication commitment to provide this information. This may suggest that editors, referees and readers are confident that the empirical results presented in the papers are always credible and robust. However, this is not always the case (Lacetera and Zirulia, 2011). In their seminal paper, Dewald et al. (1986) attempted to replicate 54 papers published in the Journal of Money, Credit and Banking and could replicate only two. Later, McCullough et al. (2006) tried to replicate 69 articles published in the same journal and could only replicate 14. McCullough et al. (2008) attempted to replicate 117 articles published in the Federal Reserve Bank of St. Louis Review and could only replicate 9. These findings raise serious concerns regarding the credibility and reliability of empirical work.[5]

We contribute to this first strand of literature in three important aspects. First, we provide *systematic* empirical evidence on the status quo in economics and management with respect to voluntary data-sharing. Second, we find strong empirical support for the hypothesis that personal attitudes towards open science principles drive the choice of researchers to share data. Third, to the best of our knowledge, this is the first paper to empirically test (and find empirical support for) the hypothesis that mandatory data-disclosure induces voluntary data-sharing.

Finally, this paper relates to the above-mentioned second strand of literature on the incentive structure of sharing of intermediate research results. Haeussler et al. (2014) find that tenured academics in the life sciences are more likely to engage in information-sharing in one-on-one situations, e.g., cases in which one researcher requests specific material or data from another. However, they also find that tenure does not matter in the case of general sharing, e.g., presentations of intermediate research results at conferences or workshops. Our analysis differs from Haeussler et al. (2014) in two aspects. First, we find that tenure can also matter for data-sharing via author websites or public data repositories. Second, while Haeussler et al. (2014) adopt a survey response, we derive our results from observation of the websites of randomly selected researchers and public data repositories.

The remainder of the paper is organised as follows. In Section 2, we describe the conceptual framework and develop our research hypotheses. Section 3 presents the data and defines the variables

[2] While our preferred definition of voluntary data-sharing does not comprise mandatory data-disclosure through journals' data archives, we discuss its possible inclusion as part of voluntary data-sharing in Section 4.2.

[3] Most notably, Carmen M. Reinhart and Kenneth S. Rogoff *did* provide their research data and, thereby, facilitated the detection of their coding error.

[4] Note that, at least in terms of transparency of published results, theoretical econom(etr)ics and applied economics appear to follow different approaches. In theoretical economics, the process by which authors obtain their results is typically transparent and theorems can be (and frequently are) checked by either the referee or the interested reader (Anderson et al., 2008).

[5] This phenomenon is not limited to economics and management. In medicine, for example, replication of published preclinical studies often fails (Begley and Ellis, 2012) and most authors do not fully adhere to data availability policies of high-impact journals (Alsheikh-Ali et al., 2011).

under study. In Section 4, we run the regressions and report the results. Section 5 discusses resulting policy implications and Section 6 concludes our study.

2. Conceptual framework and research hypotheses

A researcher's decision to share data depends on the incentive structure of the priority-based reward system of science such as career incentives and scientific competition (Cohen and Walsh, 2008; Feigenbaum and Levy, 1993; Haeussler, 2011; Haeussler et al., 2014; Stephan, 1996). It also depends on institutional factors, e.g., data availability policies of journals (Anderson et al., 2008; McCullough, 2009) and on personal attitudes towards open science (Dasgupta and David, 1994; Mukherjee and Stern, 2009). In this section, we review the existing literature on voluntary data-sharing and mandatory data-disclosure and derive five testable hypotheses.

We distinguish between three groups of testable hypotheses: first, personal incentives to share, second, institutional factors and, third, personal attitudes towards open science. The first relates to personal incentives of the researchers who make a decision (not) to share their data. In contrast, the second type of hypotheses addresses the environment researchers are working in. For instance, it considers whether the universities researchers are affiliated with facilitate data-sharing through, e.g., data management services and whether the journals in which researchers publish have data availability policies. The third type addresses heterogeneity with respect to personal attitudes towards open science among researchers.

2.1. Incentives

In a theoretical model of article publication and the decision to share data, Feigenbaum and Levy (1993) suggest that researchers without academic tenure have a higher incentive than tenured professors to keep their data secret. The intuition behind this theoretical result is the following. First, data-sharing is less beneficial for untenured researchers since the value of an additional published article is higher for untenured researchers than for tenured ones. The present value of an additional published article decreases over the lifecycle of a researcher (Coupé et al., 2006; Siegfried and White, 1973; Stephan, 1996; Tuckman and Leahey, 1973). By sharing their data, researchers create competition for themselves and reduce the competitive advantage associated with the data (Haeussler, 2011; Haeussler et al., 2014; McCullough, 2009). This competitive advantage is particularly valuable for researchers who need additional publications to get tenure. Hence, researchers who compete for tenure are less likely to voluntarily give up their competitive advantage of unique data.

Second, data-disclosure reduces the cost to peers of checking published results (Lacetera and Zirulia, 2011). Tenured researchers who already have substantial publishing experience might be less likely to commit errors than less experienced and untenured researchers (Feigenbaum and Levy, 1993). Hence, untenured researchers may have a higher incentive to hide their lack of experience (or even carelessness) by not making data available (Feigenbaum and Levy, 1993).

Third, as tenure in economics and management is typically obtained by publishing rather than by (positive) replication of one's work, the (untenured) researcher does not have an incentive to facilitate replication of her works by sharing her data.[6] On the basis of the above-mentioned arguments, our first hypothesis is as follows:

Hypothesis 1. Tenured researchers are more likely to share data voluntarily.

Our second hypothesis addresses the relation between the quality of researchers and their decision to share data (Anderson et al., 2008). Data-disclosure not only reveals information on research methods and results, but also on the quality of the researcher (Dasgupta and David, 1994). Higher-quality researchers are (ceteris paribus) more likely to endorse the principles of open science to signal their quality. For instance, higher-quality researchers are more likely to produce robust results (Feigenbaum and Levy, 1993). In addition, unknown researchers of lower quality are under higher scrutiny from their peers as they perceive them as less credible than well-known high-quality researchers (Lacetera and Zirulia, 2011). Finally, higher-quality researchers may be more likely to share data because they can more efficiently extract the full personal value of a data set with the first publication and are more likely to have a larger stock of promising research projects in the publication pipeline. Hence, the value of subsequent publications using the same data might be lower for higher-quality authors. Given these arguments, we propose:

Hypothesis 2. Better researchers are more likely to share data voluntarily.

2.2. Institutional factors of data-sharing

A second set of hypotheses relates to the institutional factors of journals and universities that influence data-sharing (Tenopir et al., 2011). First, top-tier economics journals have recently introduced data availability policies (Anderson et al., 2008; McCullough, 2009). In general, three major types of journals can be distinguished: the first has no requirement to share; the second requires (or encourages)[7] data-sharing upon request and the third type has mandatory data-disclosure after acceptance but prior to publication. The creation, documentation and transformation of data so as to increase the resilience of empirical research, which can then be made publicly available under such data availability policies comes at a cost in terms of effort and time (Feigenbaum and Levy, 1993; McCullough et al., 2006). Once the time and effort have been invested in order to meet the data-disclosure requirement of a journal, these costs are sunk. Thus, researchers might be more inclined to (also) share voluntarily. For instance, it would take virtually no time or effort for researchers to set a link on their websites to the data archive of a journal where their data are already posted. Moreover, if the data has been published already in the context of one journal publication, not disclosing the data in another journal that does not have a mandatory data-disclosure policy or on the author's website does no longer entail advantages, since benefits from non-disclosure are foregone. On the basis of the above-mentioned arguments, our third hypothesis is as follows:

Hypothesis 3. Mandatory data-disclosure induces voluntary data-sharing.

Our fourth hypothesis addresses the relation between institutional data policies and the data-sharing behaviour of affiliated

[6] However, even if positive replication would matter more for the tenure decision, economic theory suggests that the economic profession is currently in ä low-replicability equilibrium̈(Anderson et al., 2008, p. 113).

[7] The Crawford and Sobel (1982) model can typically be applied to situations where keeping one's private information (partially) secret influences the decision by another agent. Following Anderson et al. (2008), the interaction between an author and a data-requesting journal editor in the case of encouraged sharing exhibits characteristics of the problem of strategic information transmission in the classical sense of Crawford and Sobel (1982).

researchers. Building on Scott's (2001) institutional theory[8] and Ajzen's (1991) theory of planned behaviour, Kim and Stanton (2013) analyse the institutional factors driving researchers' data-sharing behaviour. They suggest that perceived effort costs of compilation, preparation and sharing research data negatively influence researchers' data-sharing behaviour, whereas normative pressure facilitates data-sharing. Universities may facilitate data-sharing of affiliated researchers by offering structured support for the creation of data management plans or statements, which are now frequently required by funding agencies (NIH, 2003; NSF, 2011). In addition, they may facilitate data-sharing by establishing data retention policies and institutional data repositories (Kim and Stanton, 2013). These measures reduce the (perceived) effort cost of the individual researcher to comply with costly external data-sharing requirements and support the adequate compilation, preparation and sharing of research data.[9] They may also shape a normative institutional environment that guides the behaviour of affiliated researchers towards data-sharing (Kim and Stanton, 2012, 2013).[10] Inspired by the Prisoner's Dilemma game of information sharing in science (Dasgupta and David, 1994; Haeussler et al., 2014), we finally argue that institutional policy provides an instrument to move away from the non-cooperative equilibrium of non-disclosure of data. Given these arguments, we propose the following hypothesis.

Hypothesis 4. Researchers from universities with institutional policies that facilitate data-sharing are more likely to share data voluntarily.

2.3. Personal attitudes towards open science principles

Finally, data-sharing might be related to personal characteristics of individuals (Constant et al., 1994; Wang and Noe, 2010). For instance, some researchers might have personal attitudes towards open science principles (Merton, 1973; Mukherjee and Stern, 2009). Open science principles place normative value on the sharing of research results and might also play an important role in the decision to share data. Inspired by Stern (2004), a recent strand of literature on the willingness of scientists (in industrial R&D departments) to pay to engage in open science suggests that there is a significant heterogeneity with respect to scientists' attitudes towards openness and their "taste for science" (Agarwal and Ohyama, 2013; Roach and Sauermann, 2010; Sauermann and Roach, 2014). Sauermann and Roach (2014) find that a stronger taste for science, e.g., for contributing to the expansion of public knowledge, significantly increases the personal value of publishing. Agarwal and Ohyama (2013) and Roach and Sauermann (2010) observe a significant heterogeneity of scientists with respect to their taste for nonmonetary returns in terms of reputation and recognition in an environment of "openness and sharing" (Roach and Sauermann, 2010, p. 423). Another strand of literature suggests that researchers are heterogeneous in terms of their attitudes towards open access publishing (Mann et al., 2009; Warlick and Vaughan, 2007; Xia, 2010).[11] Finally, many scientists may feel repugnance towards attempts to limit access to knowledge for private benefit or profit (Gans and Stern, 2010). In the light of these findings, we propose that personal attitudes towards open science are another potential factor – distinguishable from personal career incentives and institutional factors – that drives voluntary data-sharing.

While such personal attitudes towards openness are unobserved, we argue that they are likely to be highly correlated with other types of sharing. Hence, we collect data on whether the researchers under study share other material and take this as a proxy for personal attitudes towards open science. We propose:

Hypothesis 5. Personal attitudes towards open science affect voluntary data-sharing.

3. Data and definition of variables

In our data set we gather information on the extent of voluntary data-sharing and mandatory data-disclosure of 488 empirical researchers in economics, 388 of which are affiliated with an economics department (100 with a business school). The researchers are chosen uniformly across the top 100 economics departments (four observations each) and top 50 business schools[12] (two observations each) and randomly within the respective institution.[13] Every researcher in the sample is an empirical economist. These researchers are selected based on their sub-discipline, research interests and courses taught. The keywords we consistently look for are: "applied" and "empirical". The economics departments and business schools under study typically provide overviews of the current position of their faculty, i.e., whether the researcher has tenure or not, on their websites where they indicate the chairs, sub-disciplines and research interests of affiliated researchers. In addition, the researchers typically indicate their research interests, courses taught and journal publications on their websites and CVs. Using these sources of information, we create a complete list of researchers who fulfil our selection criteria for each institution. From the list of potential candidates, we randomly select[14] four (two) researchers from economics departments (business schools). We also double-check the empirical research focus of the researchers under study by looking at the keywords of their articles published in journals that are listed in the Web of Science (WoS).

The main variable of interest is whether researchers share their data in a transparent and easily accessible way. Currently, there is no universally accepted standard procedure through which data ought to be shared. Researchers can choose among many

[8] Institutional theory explores the extent to which the institutional environment drives individual action through regulative, normative and cultural-cognitive channels (Currie and Suhomlinova, 2006; DiMaggio and Powell, 1991).

[9] See also Furman and Stern (2011) for a recent analysis of the impact institutions have on knowledge diffusion and knowledge accumulation when they lower the costs of access to knowledge.

[10] On a more general level and inspired by Merton (1973), Rhoten and Powell (2007) and Haeussler (2011) suggests that a scientific community's adherence to the norms of open science influences the individual sharing behaviour of its members. More specifically, the greater are researchers' beliefs that the norms of open science operate in their community, the greater is the expected level of general information-sharing (Haeussler et al., 2014).

[11] See Mueller-Langer and Scheufen (2013) for a recent overview of the literature on academic publishing and open access. See also McCabe and Snyder (2005), Suber (2012) and Mueller-Langer and Watt (2010, 2014).

[12] All selected researchers affiliated with business schools have an economics background and were selected from among researchers who taught economics-related courses, e.g., applied microeconomics, at business schools. The group of business-school economists was selected in this particular manner to make it as comparable as possible with the group of economists working in economics departments.

[13] Three business schools (London Business School, INSEAD and HEC Paris) are listed in both rankings. Thus, we exclude 12 observations from the top 100 economics departments in our data set.

[14] For example, if researchers teach applied econometrics, empirical finance, econometric evaluation, applied macroeconomics or applied microeconomics, we would categorise them as ëmpiricaländ thus eligible for the random draw. In contrast, if their primary research interests are economic theory or theoretical game theory, we would exclude them from the list of possible candidates per institution. In addition, the publication lists retrieved from Thomson Reuters Web of Science provide valuable information on the fields of interest and research areas of the researchers under study. Researchers who systematically publish in journals such as Ëconomic Theoryäre excluded while researchers publishing in journals like Äpplied Economicsäre included in the list of empirical researchers.

Please cite this article in press as: Andreoli-Versbach, P., Mueller-Langer, F., Open access to data: An ideal professed but not practised. Res. Policy (2014), http://dx.doi.org/10.1016/j.respol.2014.04.008

channels of data-sharing such as their personal or institutional websites, public data repositories or data archives of journals. In the following sections, we distinguish between two types of data availability: voluntary data-sharing and mandatory data-disclosure. We first consider voluntary data-sharing via researchers' personal or institutional websites and public data repositories. Then, we analyse voluntary data-sharing as a consequence of mandatory data-disclosure policies. The key distinction is that voluntary data-sharing is a choice by researchers, while mandatory data-disclosure policies impose a constraint on authors.[15] In the empirical analysis, we compare the regression coefficients using these two dependent variables to evaluate whether voluntary data-sharing exhibits a systematically different relation to observable researchers' characteristics than mandatory data-disclosure.

Table 1 provides an overview of the dependent and independent variables under study, which we explain in more detail in the following sections.

3.1. Dependent variables

3.1.1. Data-sharing as voluntary strategy of researchers (model 1)

The dependent variable in our first regression model is *VolDataSharing*, an ordinal variable that represents the extent to which researchers voluntarily make their data available in a "*clearly and precisely documented*" way and "*readily available to any researcher*" (AER, 2013).[16] For this purpose, we focus on researchers' institutional or personal websites and data entries of the researchers under study in public data repositories. We search the websites of the 488 researchers under study for a "data and code" section or, on a paper-by-paper-basis, links to public data repositories or journal data archives. In addition, we examine whether the researchers under study have entries in public data repositories.[17] In particular, we search for data and code entries in the Harvard Dataverse Network, the Inter-University Consortium for Political and Social Research (ICPSR) Data Archive, the Institution for Social and Policy Studies (ISPS) Data Archive (Yale University) and RunMyCode. For instance, 43 of the researchers under study (8.81%) have, in total, 105 data and code entries for scientific articles in Dataverse. Dataverse is a self-service repository where individual researchers, journals and institutions can create their own dataverses. For instance, the Network of European Economists Online has its own dataverse and is included in our analysis (27 entries). Additional institutions with own dataverses included in our analysis are the Abdul Latif Jameel Poverty Action Lab (three entries) and the Better Access to Data for Global Interdisciplinary Research Initiative (one entry). Interestingly, 23 (21.9%) of the total entries on Dataverse are from the one researcher who has his own dataverse (Joshua Angrist). Eleven of the researchers under study (2.25%) have, in total, 30 data and code entries for scientific articles in the ICPSR Data Archive.[18] Five of the researchers under study (1.02%) have, in total, five data and code entries for scientific articles on RunMyCode. Finally, one of the researchers under study (0.2%) has five data and code entries for scientific articles in the ISPS Data Archive.

VolDataSharing takes on three values: no sharing (0), sporadic sharing (1), and full sharing (2). The values of sharing are defined as follows: *VolDataSharing* = 0 indicates that researchers do not share *any* of their data through their websites or public data repositories and do not have a link or any other type of reference as to whether the data is available. Notably, 394 of the researchers under study (80.74%) fall within this category.[19] *VolDataSharing* = 1 indicates that researchers share their data in a sporadic manner. Some isolated data and codes or a link to a public data repository or journal data archive can be found but for most papers neither the data nor the code is accessible or clearly described. In our sample, 82 researchers (16.8%) fall under this category. Finally, *VolDataSharing* = 2 indicates that researchers have a well-designed, clear and freely accessible data and code section on their websites or share additional information on the availability of most of the data they use in public data repositories. In our sample, only 12 researchers (2.46%) fall under this category. Our definition of the category "full sharing" requires that a large amount, though not necessarily all, of the data is available. Data-sharing is not generally considered a norm in economics and there is no clear and widely accepted definition of how an "ideal" researcher's data archive should be designed.[20] We define the data-sharing behaviour of an author as *VolDataSharing* = 2 if she makes data available for at least 50% of her empirical papers. In addition to sharing not being the norm, there are two other main reasons we set this threshold at 50%: first, some data sets cannot be made freely available due to privacy issues or because they are proprietary. Second, data from old papers written before the data-sharing debate started might be either lost or in old formats that are difficult to convert. In short, *VolDataSharing* is designed to reflect the heterogeneity in the observed sharing decision by researchers and does not reflect any ideal standard of sharing.

3.1.2. Mandatory data-disclosure as institutional built-in practice (model 2)

In contrast to the previous section, we here consider mandatory data-disclosure as an institutional built-in practice of journals with data availability policies. More specifically, we analyse data policies of the top 100 (50) economics (business) journals according to the 2012 impact factor rank provided by the Thomson Reuters Journal Citations Reports (Social Sciences Edition, as of August 2013).[21]

In order to identify articles that are subject to a data availability policy, we first identify journals that have such a policy in place as well as the first issue in which the policy was adopted. The description of the data availability policy is often publicly available, i.e., in the author guidelines or in a separate section on the journals' websites. In contrast, the exact date of the introduction of the data availability policy is not available online in the vast majority of the

[15] Clearly, authors decide where to submit their papers with full awareness of data availability policies. However, they still have the additional choice of whether to actively promote replication by having a data and code section or referring to the journals' data archives on their websites.

[16] *VolDataSharing* captures two (indistinguishable) types of sharing. First, researchers may share their data because they believe it is part of publishing a paper, which can be referred to as pro-active data-sharing. Second, researchers may share their data on their webpages and/or public data repositories after having received a request to share by their peers. In this case, they may decide to also make the data available online as they do not incur additional effort cost by doing so.

[17] The public data repositories under study are selected after careful consideration of the suggestions of the IT specialists of the European Data Watch Extended Project and a thorough search on the Internet. Our main selection criteria are the number of entries and the relevance for economics and management.

[18] In contrast to Dataverse, the available data on the curated ICPSR Data Archive is often government data that is required to be publicly available.

[19] Note that a link to the journal data archive where the paper and the data used are published would be sufficient to exclude a researcher from the non-sharing category.

[20] While one might argue that all the data and codes of published papers should be made available, others have contrasted this view by arguing that sharing would discourage researchers from creating data as after the first publication they would not be the full residual claimant anymore. This might lead to under-provision of empirical papers (Moffitt, 2007).

[21] To avoid double entries, we exclude the Journal of Business Economics and Management and the Journal of Environmental Economics and Management from the list of the top 100 economics journals as they are listed under both WoS research areas economics and business. To avoid double entries, we also exclude IMF Staff Papers from the top 100 economics journals list, as it has ceased publication and was replaced by IMF Economic Review in 2010.

Table 1
Summary statistics.

	Mean	Median	St. dev.	Min	Max	Obs.
Voluntary data-sharing	0.217	0	0.469	0	2	488
Mandatory data-disclosure	0.103	0.0137	0.170	0	1	488
Share other material	0.186	0		0	1	488
Tenure	0.529	1		0	1	488
Log average citations	2.475	2.575	1.209	0	6.070	488
Business school	0.205	0		0	1	488
Inst. data policies	1.381	1	0.887	0	3	488
% Top-5 papers (econ.)	0.124	0.0377	0.181	0	1	488
Experience	17.96	16	12.27	0	50	488
Experience squared	472.9	256	551.2	0	2500	488
No. of research grants	6.717	5	8.060	0	47	488
Male	0.801	1		0	1	488
USA empl.	0.678	1		0	1	488
EU empl.	0.168	0		0	1	488

cases.[22] Therefore, we ask the editors of 33 journals for which we find preliminary evidence online for the existence of a data policy to provide us with information on the issue in which the policy was first adopted. The response rate for our email survey is 81.81%. For the remaining six journals without reply from the editors we use the issue in which data is posted on the website of the respective journal for the first time as a proxy for the introduction of the policy.[23]

Twenty journals (13.6%) in our sample have a mandatory data-disclosure policy which typically follows the mandatory data-disclosure policy of the American Economic Review, which states the following (AER, 2013): "*It is the policy of the American Economic Review to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Authors of accepted papers that contain empirical work, simulations, or experimental work must provide to the Review, prior to publication, the data, programs, and other details of the computations sufficient to permit replication. These will be posted on the AER Web site*". Ten journals (6.8%) require (in some cases, encourage) data-sharing upon request by interested researchers but data are typically not posted in the journals' websites. 117 journals (79.6%) have no data availability policy. Henceforth, we distinguish between the 30 journals with a data availability policy (mandatory data-disclosure and required (encouraged) data-sharing upon request) and the 117 journals without any data availability policy.

Using this information, we construct a variable that indicates the percentage of a particular author's articles that are subject to a data availability policy at the time of publication.[24] On average, 10.3% of an author's publications are subject to a data availability policy. This variable captures the strength of the obligation on authors to share their data, whereas *VolDataSharing* reflects the choice to share. As the obligation to share might also induce authors to share voluntarily, we will use it both as an explanatory variable in the first regression model and as a dependent variable in the second regression model.

3.2. Independent variables

The independent variables are split into four categories: personal incentives, institutional factors, personal attitudes towards open science and control variables.

3.2.1. Incentives

We first describe the variables that relate to personal incentives to share: *Tenure* and *LogAverageCitations*. *Tenure* is a binary variable indicating whether or not the researcher is a tenured professor. This information is collected by looking at the CVs of researchers and the departments' websites. Of the researchers under study, 258 (52.87%) are tenured professors. The second variable, *LogAverageCitations*, is a proxy for the ability of researchers and how highly they are esteemed by their peers. For the 488 authors under study, we retrieve article meta-data for their complete set of articles published in WoS-listed journals from 1958 to 2013.[25] In particular, we gather information on the title, list of authors, journal, year, volume and issue for a total of 10014 articles. Using WoS-listed journals, *LogAverageCitations* is defined as the logarithm of the total number of citations by a researcher divided by her total number of papers.[26]

3.2.2. Institutional factors

In addition to personal incentives, institutional factors might affect data-sharing behaviour. The first and most important factor is the data availability policy of the journals in which researchers publish their articles. This variable is identical to the one we use as dependent variable in regression model 2. It takes on values between 0 and 1 and is defined as the percentage of papers that are subject to a data availability policy of journals at the time they are published. The second variable that relates to institutional factors reflects the number of data management tools and services a university provides for its researchers. This variable can be thought of as the extent to which a university has policies to facilitate data-sharing. *InstDataPolicy* takes on four values, from 0 (no policy facilitating data-sharing) to 3 (full set of policies). It is defined as the sum of three dummy variables[27]:

[22] In a few exceptional cases, editors provide information and data availability policies and their date of introduction in editorial notes or reports.

[23] These journals are Value in Health, Applied Economic Perspectives and Policy and the Australian Journal of Agricultural and Resource Economics in economics and the Journal of Marketing, the Journal of the Academy of Marketing Science and the Journal of International Marketing in management.

[24] For example, if an author has ten AER papers, of which seven are published before the mandatory data-disclosure policy was first adopted, this variable will take on the value 0.3.

[25] Mueller-Langer et al. (2013) describe the exact process of the creation of author-specific citation metrics and publication data that is adopted in the present article.

[26] As some researchers have no citations, we define *LogAverageCitations* = log(AverageCitations + 1). We also generate other proxies for ability such as the h-index. However, as we will discuss in Section 4, all measures yield comparable results.

[27] We use this proxy for the existence of institutional data-disclosure policies, as the institutions under study do not have mandatory data-disclosure policies. Notably, we find explicit data-disclosure policy statements from the University of Pittsburgh (2009) and Weill Cornell Medical College (2007). However, these policy statements specifically consider the accessibility of medical research data only.

OwnResearchDataRepository indicates whether the institution under study runs its own data repository that makes research data of affiliated researchers publicly available. *DataManagementPlan* indicates whether the institution provides guidance for the creation of data management plans, e.g., through online tools, manuals or workshops.[28] *ResearchDataRetentionPolicy* indicates whether the institution has an explicit research data retention policy for affiliated researchers.

Finally, we generate a dummy variable to control for data-sharing differences between economics departments and business schools. *BusinessSchool* is a binary variable indicating whether the researcher works in a business school or in an economics department.

3.2.3. Personal attitudes towards open science

Given the same incentives, people with different social and cultural norms tend to respond differently. In our context researchers faced with the same incentive structure might choose to share because of social and cultural norms with respect to sharing rather than because of the incentives. Such norms go beyond sharing final research results and encompass sharing lecture material such as slides and taking part in open knowledge projects. Not accounting for such heterogeneity might bias the regression results. Some researchers might feel obliged (or even intrinsically motivated) to share data for replication purposes while others do not. These personal attitudes are unobserved and, thus, we construct a proxy for them. *ShareOtherMaterial* is a binary variable that indicates whether researchers share other material, such as lecture slides, on their websites. Researchers are typically not forced to share the syllabus of the classes they teach or the lecture slides to students not enrolled in the course but some lecturers do this anyhow. We argue that sharing such material provides us with a good proxy for personal attitudes towards principles of openness in general.

3.2.4. Controls

Finally, we include a set of control variables. *Experience* measures how long the researcher has been writing papers and is defined as the number of years since earning a Ph.D. In addition, we include the square of *Experience, ExperienceSquare*, to control for quadratic effects. To control for top publications we use researchers' share of publications in the top-five economics journals according to Card and DellaVigna (2013) in their total publications as an additional measure of the ability of authors.[29] The 488 authors in our sample published a total of 1388 articles in the top-five economics journals (out of a total of 10014 articles).[30] *ResearchGrants* indicates the total number of research grants won by the researcher which is an additional proxy for ability and resources. We hand-collect this information from the CVs of the authors under study. Finally, we include *Male*, *USAempl* and *EUempl*, which are binary variables equal to 1 if the researcher is male, works in the U.S. or Europe, respectively.

4. Empirical results

This section presents and discusses the empirical evidence with respect to the two aims of this paper: first, to examine the status quo of data-sharing by empirical researchers in economics and management and, second, to analyse the extent to which voluntary data-sharing is driven by personal incentives, institutional factors and personal attitudes towards open science principles as discussed in the hypothesis section.

4.1. Status quo of voluntary data-sharing

The analysis of our hand-collected data provides a clear result about the current extent to which researchers voluntarily share their research data: The great majority of researchers do not share. Out of a sample of 488 empirical researchers, only 82 (16.8%) engage in some voluntary data-sharing via their websites or public data repositories. Only 12 (2.46%) provide data and codes in a clearly and precisely documented way, readily available to any researcher. In contrast, the large majority, 394 (80.74%) researchers, neither provide the data nor indicate whether or where the data used in their empirical work are available. This result is in line with prior literature on data-sharing, which emphasizes that economists *demand* data-sharing but are not willing to *provide* it (Anderson et al., 2008; Hamermesh, 2006; McCullough et al., 2006).

4.2. Main regression model: voluntary data-sharing

This section presents the empirical results of our main regression model, which tests the hypotheses that we develop in Section 2. We distinguish between three sets of hypotheses that relate to personal incentives, institutional factors and personal attitudes towards open science.

To test these hypotheses we run an ordered probit model where the dependent variable is *VolDataSharing*, an ordinal variable that reflects the extent to which researchers voluntarily make their data available on their websites or public data repositories. It takes on three values, where 0 indicates no sharing, 1 sporadic sharing and 2 full sharing as discussed in section 3. Table 2 presents our main regression results on the determinants of voluntary data-sharing. As we move from specification (1)–(3) we add more control variables, while in specification (4) we run our full model without *ShareOtherMaterial*. This final specification tests whether adding *ShareOtherMaterial*, i.e., the proxy for personal attitudes towards open science, significantly affects the other coefficients. Because the coefficients do not change significantly once *ShareOtherMaterial* is included, we focus on specification (3) to compute the marginal effects presented below.

Consistent with the hypotheses derived in Section 2, our main variables of interest are: *Tenure*, *LogAverageCitations*, *MandatoryDataDisclosure*, *InstDataPolicy* and *ShareOtherMaterial*. In addition, we include a set of variables to control for potential sources of unobserved heterogeneity. *Experience* and *ExperienceSquare* control for the number of years the researcher has been doing research. This is an important control with respect to the effect of tenure (see Hypothesis 1). To test whether tenured professors share more we need to isolate the career effect from the number of years spent doing research. *ResearchGrants* accounts for possible differences in resources across researchers. *Male* is included to control for gender-related differences in sharing behaviour.[31] *USAempl* and *EUempl* control for location effects. The share of top-five papers in economics journals is an additional control for the quality of the researcher. It also accounts for the fact that four of the top-five

[28] We observe that only few European universities provide this type of guidance. This finding can be explained by the fact that it is mainly American research-funding institutions that push for the establishment of data management plans (NIH, 2003; NSF, 2011).

[29] The top-five economics journals are AER, Econometrica, Quarterly Journal of Economics, Journal of Political Economy and Review of Economic Studies.

[30] We also tried a specification including the author's share of publications in top-5 management journals (results not shown). However, as all the economists in the sample tend to publish in the top-ranked economics journals, the share of top-5 management publications is insignificant and does not change the results.

[31] Twelve female researchers (out of a total of 97) in our data set (partly) share data. Thus, we do not have sufficient data to study gender effects and the coefficient on *Male* should be interpreted accordingly.

Table 2
Voluntary data-sharing.

	(1)	(2)	(3)	(4)
Dependent variable:	*VolDataSharing*: ordinal; 0 = no sharing; 1 = sporadic sharing; 2 = full sharing			
Regression model:	Ordered probit			
Tenure	0.407**	0.574***	0.569***	0.554***
	(0.171)	(0.205)	(0.213)	(0.211)
Log average citations	0.240***	0.312***	0.298***	0.273***
	(0.0706)	(0.0832)	(0.0838)	(0.0822)
Mandatory data-disclosure	1.611***	1.811***	1.779***	1.597***
	(0.395)	(0.543)	(0.541)	(0.531)
Business school	−0.0605	−0.0534	−0.0673	−0.103
	(0.183)	(0.185)	(0.187)	(0.184)
Inst. data policy	0.106	0.120	0.0730	0.0823
	(0.0786)	(0.0813)	(0.0931)	(0.0918)
Share other material	0.530***	0.567***	0.548***	
	(0.157)	(0.160)	(0.162)	
% Top-5 papers (econ.)		−0.496	−0.423	−0.148
		(0.517)	(0.521)	(0.506)
Experience		−0.0357	−0.0341	−0.0242
		(0.0275)	(0.0278)	(0.0274)
Experience squared		0.000556	0.000495	0.000269
		(0.000520)	(0.000525)	(0.000515)
No. of research grants		0.000358	0.000706	0.000306
		(0.00826)	(0.00833)	(0.00824)
Male			0.212	0.283
			(0.201)	(0.199)
USA empl.			0.288	0.293
			(0.252)	(0.248)
EU empl.			0.232	0.200
			(0.274)	(0.270)
Constant (cut1)	2.202***	2.085***	2.397***	2.350***
	(0.232)	(0.282)	(0.359)	(0.357)
Constant (cut2)	3.463***	3.356***	3.674***	3.589***
	(0.275)	(0.317)	(0.388)	(0.384)
Observations	488	488	488	488

Note: The dependent variable in all six specifications is *VolDataSharing*, which is ordinal and takes on three values. All coefficients are estimated using an ordered probit model. The marginal effects reported in the text have been computed using specification (3). Standard errors are reported in parentheses.
* Significance level is reported as $p < 0.1$.
** Significance level is reported as: $p < 0.05$.
*** Significance level is reported as $p < 0.01$.

economics journals have a mandatory data-disclosure policy.[32] Controlling for the share of top-five publications is especially important to evaluate the effect of *MandatoryDataDisclosure* on voluntary data-sharing, as the two variables are positively correlated. Finally, *BusinessSchool* allows us to control for institutional differences between economics departments and business schools with respect to data-sharing. Table 2 shows that none of these control variables is significant throughout the specifications. However, we believe that it is still important to include them in the model.[33] As discussed above, we consider specification 3 to be our preferred model and we compute the marginal effects using this specification.

4.2.1. Hypotheses 1 and 2: incentives

The first hypothesis we test is whether tenured researchers are significantly more likely to share their data. The coefficient on *Tenure* is positive and highly significant at the 1% [5%] level in specification (2), (3) and (4) [(1)] as reported in Table 2. Tenure increases the likelihood of sporadic sharing (full sharing) as compared to no sharing (sporadic sharing) by 11.7% (1.7%) using the full model in specification (3). This finding provides empirical support for Hypothesis 1. The career concerns of junior researchers seem to play an important role in the decision to share data. Data-sharing is a form of self-induced competition, as it permits other researchers to use a data set before its creator can fully exploit it in further research. An additional publication is arguably more valuable in terms of career planning for (untenured) junior researchers than for tenured professors. In addition, junior researchers might be more likely to commit errors than more experienced, tenured professors, which would increase the potential drawbacks of data-sharing for the former.

Hypothesis 2 considers the relation between the quality of researchers and their decision to share. While quality is unobserved, there are several proxies we can use, e.g., average citations per paper, sum of citations per author or the h-index. In general, citation metrics are a commonly accepted indicator for the success of researchers and the extent to which peers value their work. We use each of the three citation metrics. However, as they yield similar results, we focus on a transformation of average citations, i.e., the logarithm of average citations per paper, in order to account for the skewedness of average citations. *LogAverageCitations* is positive and highly significant at the 1% level across all four specifications. A unit increase of *LogAverageCitations* increases the likelihood of sporadic sharing (full sharing) as compared to no sharing (sporadic sharing) by 6.3% (0.9%). This effect is highly significant and thus confirms the theoretical predictions derived in

[32] Note that none of the top-five management journals (according to the 2012 impact factor rank provided by Thomson Reuters Journal Citations Reports) has so far introduced a mandatory data-disclosure policy. We discuss discipline-specific differences with respect to mandatory data-disclosure policies in Section 5 (Policy Implications).

[33] This choice does not have a significant effect on the magnitude of the coefficients of interest, while including these controls appears to be reasonable from a theoretical perspective.

Hypothesis 2. Higher-quality researchers are more likely to voluntarily share data.[34] Based on Dasgupta and David (1994), we argue that higher-quality researchers share their data in order to signal their quality. They may also have many original research projects in the publication pipeline at any given time and, thus, attribute a low value to an additional publication using the same data.

4.2.2. Hypotheses 3 and 4: institutional factors

We now focus on the second set of hypotheses related to institutional factors. In contrast to personal incentives, institutional factors relate to the environment a researcher is exposed to. The major change in recent years has been the introduction of mandatory data-disclosure policies by some journals. These policies might play a major role in the future of data-sharing and, most importantly, induce more voluntary data-sharing as discussed in Section 2.2. We find that *MandatoryDataDisclosure* is positive and highly significant at the 1% level in all four specifications.[35] An increase of one standard deviation in the share of papers published under a data policy, i.e., 17%, increases the likelihood of sporadic sharing (full sharing) as compared to no sharing (sporadic sharing) by 6.4% (0.9%). Once the time and effort cost of preparing the data are sunk and the benefits from non-disclosure are foregone due to mandatory data-disclosure, the likelihood of voluntary data-sharing increases. In addition, mandatory data-disclosure policies might also change how data-sharing is regarded by the community. If most journals adopt such a policy, the patent view of data described by Moffitt (2007) might shift towards a new norm under which data-disclosure is an inseparable part of a publication.[36] While one might argue that better or more senior researchers choose to publish in journals with a data availability policy, the estimated marginal probabilities are conditional on *LogAverageCitations*, *Tenure*, *%Top5Publications*, *ExperienceSquare* and *Experience*. These results suggest that mandatory data-disclosure policies have a considerable impact on how sharing is perceived and lead to more voluntary data-sharing.

A second institutional factor affecting voluntary data-sharing is the set of data-sharing-related policies of the universities with which researchers are affiliated. We hypothesise that researchers at universities with a higher availability of data-sharing tools and services are more likely to share (Hypothesis 4). Thus, we expect the coefficient on *InstDataPolicy* to be positive. In fact, we find that *InstDataPolicy* is positive in all specifications but not significantly so. Our results confirm that the relation goes in the expected direction. However, this relation is statistically weak and the magnitude is not large.

4.2.3. Hypothesis 5: personal attitudes towards open science

Finally, Hypothesis 5 considers the researchers' personal attitudes towards open science. While personal attitudes are unobserved, they might play an important role in the decision to share data. As a proxy we use *ShareOtherMaterial*, a dummy variable which indicates whether researchers make additional material, e.g., lecture slides, available on their websites. We hypothesise that researchers who conform to the norms of open science are more likely to voluntarily share data. In fact, we find that *ShareOtherMaterial* is a strong predictor of voluntary data-sharing and highly significant at the 1% level. Sharing other material increases the likelihood of sporadic sharing (full sharing) as compared to no sharing (sporadic sharing) by 12.9% (2.4%). This result suggests that personal attitudes towards open science principles are the strongest predictor for voluntary data-sharing.[37]

4.2.4. Robustness check: mandatory data-disclosure

A potential concern of the regression results presented above refers to our definition of voluntary data-sharing. As there is no unique standard for data-sharing, different researchers might share their data in different ways. Our definition comprises two main sources of sharing: websites and public data repositories. However, the recent introduction of data availability policies and data archives by journals provides researchers with an additional channel for data-sharing. For instance, some authors might decide to share only through journal data archives. If these authors do not make any reference to the journals' data archive in their websites (although this would be virtually costless, as we argue in Section 2), they would appear as non-sharing researchers in our main variable of interest, *VolDataSharing*. Not accounting for this type of data-sharing might bias the results, as we would omit researchers who share data in a different way. To test whether our definition of voluntary data-sharing should include authors who are required to share via journal data archives but do not make any reference to these archives on their websites, we run a regression using *MandatoryDataDisclosure* as the dependent variable. The aim of this regression is to test whether mandatory data-disclosure and voluntary data-sharing are structurally different or can be explained by the same factors. If researchers do not engage in voluntary data-sharing because they rely on journal data archives, then our main dependent variable, *VolDataSharing*, might not be accounting for this type of sharing and, thus, it might be biased downward. In contrast, if our choice of defining data-sharing through the personal or institutional website and data repositories already accounts for the full extent of voluntary data-sharing, we would expect the two variables to follow a different data-generating process. In this case, mandatory data-disclosure should be driven by the incentives to publish in top journals, most of which have a mandatory data-disclosure policy, rather than by the other factors discussed above. Thus, if voluntary data-sharing is properly defined, voluntary data-sharing reflects a free choice, while mandatory data-disclosure simply reflects a constraint on researchers. Thus, in contrast to voluntary data-sharing, mandatory data-disclosure would be related to personal incentives, institutional factors and personal attitudes towards open science in a significantly different way.

We run an OLS regression using *MandatoryDataDisclosure*, the percentage of papers published by an author that were subject to a data availability policy, as the dependent variable. All explanatory and control variables are the same as in the previous

[34] However, there might be a reverse causality problem between sharing and citations. As shown by Piwowar et al. (2007) sharing research data is positively correlated with the citation rate. Authors could share their data to signal the quality and robustness of their results in order to get more citations. While this reverse effect might be present, in our regression model we control both for the share of top publications and for the experience of authors. As we go from specification (1) with no controls to specifications, (2) and (3) both the magnitude and the significance level of *LogAverageCitations* do not change significantly. If better authors used data-sharing to get more citations, the coefficient on *LogAverageCitations* should decrease once we control for author's quality. This is not supported by the regression results. In contrast, the coefficient on *LogAverageCitations* slightly increases as we go from specification (1)–(3).

[35] We try different definitions of *MandatoryDataDisclosure* following the strength of the policies, i.e., recommended data-sharing upon request or mandatory data-disclosure, but find no significant difference across definitions.

[36] One of the duties of editors is to shape the field and determine the necessary components of a paper. The radical change in data policies of some top journals and the need for replicable research might play an important role in changing the culture of data-sharing.

[37] In addition, as already argued above, the inclusion of *ShareOtherMaterial* does not significantly affect the other coefficients. This might indicate that personal attitudes towards open science are unrelated to the other explanatory variables relating to incentives and institutional policies.

model. The coefficients can now be readily interpreted as marginal probabilities.

From the results presented in Table 3 it is immediately evident that the two models are significantly different. The most important predictor of voluntary data-sharing, *ShareOtherMaterial*, is insignificant and positive in specification (1) and turns negative in specifications (2) and (3). In contrast to the results reported in Table 2, *Tenure* changes sign and has a negative and significant effect on *MandatoryDataDisclosure* in all four specifications. *LogAverageCitations* also changes sign between the first specification and the other three specifications. The key drivers of *MandatoryDataDisclosure* seem to be the negative effect of *BusinessSchool* and the high and positive effect of *%Top5Publications*, both of which are insignificant in the first regression model reported in Table 2.

These results suggest that mandatory data-disclosure is mainly driven by publication incentives rather than by any other type of personal incentives, institutional factors or personal attitudes towards open science. Four of the top-five economics journals have a mandatory data-disclosure policy and junior researchers *must* publish in these journals to advance in their careers. These researchers tend to be at the initial stage of their careers and have few publications and citations. In addition, older researchers have a lower share of publications subject to a mandatory data-disclosure policy because journals have only recently introduced such policies. This would explain the negative correlation on *Tenure* and *ExperienceSquare* and the positive correlation on *%Top5PublicationEcon*.

Finally, this evidence suggests that our main dependent variable, *VolDataSharing*, is correctly specified and indeed reflects the free choice to voluntarily share data. Voluntary data-sharing and mandatory data-disclosure are two different mechanisms and are affected by significantly different drivers. As we have shown, the latter mechanism is strongly driven by the obligation to share data imposed by mandatory data-disclosure policies. Thus, while there are convincing theoretical arguments to include mandatory data-disclosure in our main model as an explanatory variable, mandatory data-disclosure should not be included as part of the dependent variable, i.e., voluntary data-sharing, in our main regression model reported in Table 2.

5. Policy implications

The public availability of research and government data has recently attracted widespread attention from policy-makers at the national and international level (European Commission, 2012; OECD, 2007; US House of Representatives, 2007). The European Commission (2012) suggests that open access to research data will improve the quality of research results and foster scientific progress and innovation. The OECD (2007, p. 13) stresses the importance of open availability of research data to "improve the efficiency and effectiveness of the global science system". The US House of Representatives (2007) prescribes that data and results of researchers employed by a federal civilian agency must be made available to other agencies, policy-makers and the public. In addition, major national research foundations require funded researchers to publicly share research data created under their grants (Borgmann, 2012; ESRC, 2010; Haeussler et al., 2014; NIH, 2003; NSF, 2011; Wellcome Trust, 2007). Finally, academic journals have recently introduced data availability policies. However, Dewald et al. (1986) document the difficulties that journal editors face in obtaining data from authors. The reasons for these difficulties are manifold (Costello, 2009; Dewald et al., 1986; Feigenbaum and Levy, 1993; Vlaeminck et al., 2013). First, authors may not be willing to incur the cost of compiling and documenting data and codes but rather devote their efforts to producing new manuscripts. Second, they may fear jeopardising subsequent use by making data available. Finally, there is a lack of recognition for creating data and replication studies. On the basis of the above-mentioned arguments and the results presented in this paper, our policy recommendations are as follows.

5.1. Institutional factors

We find that, as of September 2013, the majority (79.6%) of the 147 top economics and management journals under study do not have a data availability policy. Twenty (13.6%) of the journals in our sample have a mandatory data-disclosure policy, of which four are among the top-five economics journals. Our results suggest that the top economics journals have taken the lead with respect to mandatory data-disclosure, whereas the lower-tier journals appear to be lagging behind. Our results also point to considerable discipline-specific differences with respect to mandatory data-disclosure policies. More specifically, 19 out of 97 economics journals under study have a mandatory data-disclosure policy, whereas only one[38] of the 50 management journals under study has such a policy. This result calls for a critical reconsideration of the data management policies of management journals.

Moreover, 6.8% of the whole journal sample require (or encourage) data-sharing upon request. They may face the above-mentioned difficulties. In particular, data availability policies that rely on encouraged data-sharing upon request rather than enforcement mechanisms often fail and, thus, are not sufficient to satisfy the ideal standard of data availability (Dewald et al., 1986; Feigenbaum and Levy, 1993; McCullough et al., 2008; McCullough and Vinod, 2003). Therefore, we recommend that economics and management journals should implement mandatory data-disclosure policies.

Finally, based on our strong finding regarding attitudes, we encourage universities and other research institutions to push for a norm change by incorporating data-sharing in their "responsible conduct of research" training and discussing its importance with Ph.D. students.

5.2. Personal incentives

In order to increase the value of data for its creator(s), we recommend that appropriate reward structures for producing and documenting data and code should be promoted (Fienberg et al., 1985), e.g., through standardised data citation (Altman and King, 2007) or by conditioning career rewards like academic tenure and research budgets on data-disclosure (Mukherjee and Stern, 2009). For instance, recently established mechanisms that aim to archive data files and make them citable are provided by Dryad, Figshare and Zenodo among others.[39] In addition, policy-makers and research-funding agencies might incentivize and support the establishment of data journals. The establishment of data journals in economics and management may create a market for the exchange of data and increase the value of a data set for its creator. This may also increase the quality of available data if data publications are subject to a peer-review process. Lessons could be learned from other disciplines. For instance, Nature (2013) has announced that it will launch an online data journal, called Scientific Data, in the spring of 2014. In addition, the Royal Meteorological Society has recently launched the online-only Geoscience Data Journal (Allan, 2012). In general, if the personal value of data as an independent scientific contribution increases due to peer-accepted data journals and data citations, researchers who have the necessary (financial)

[38] Marketing Science implemented its mandatory data-disclosure policy on 15 April 2013.

[39] See also the Open Economics Working Group's prototype of a platform for searching and storing regression results which is called MetaMetrik.

Table 3
Mandatory data-disclosure.

	(1)	(2)	(3)	(4)
Dependent variable:	*MandatoryDataDisclosure*: percentage of papers subject to a data availability policy			
Regression model:	OLS			
Tenure	−0.106***	−0.0427***	−0.0445***	−0.0439***
	(0.0171)	(0.0162)	(0.0168)	(0.0168)
Log average citations	0.0167**	−0.00961	−0.0102	−0.00995
	(0.00711)	(0.00654)	(0.00659)	(0.00660)
Business school	−0.0567***	−0.0550***	−0.0556***	−0.0536***
	(0.0184)	(0.0143)	(0.0145)	(0.0144)
Inst. data policy	0.0162*	−0.00883	−0.0103	−0.0109
	(0.00845)	(0.00669)	(0.00761)	(0.00761)
Share other material	0.00548	−0.0225	−0.0229	
	(0.0190)	(0.0148)	(0.0149)	
% Top-5 papers (econ.)		0.598***	0.601***	0.595***
		(0.0347)	(0.0351)	(0.0350)
Experience		−0.00002	0.00009	−0.000153
		(0.00215)	(0.00216)	(0.00215)
Experience squared		−0.00007*	−0.00007*	−0.00007*
		(0.00004)	(0.00004)	(0.00004)
No. of research grants		0.00009	0.000123	0.000140
		(0.000716)	(0.000721)	(0.000722)
Male			0.00883	0.00716
			(0.0149)	(0.0149)
USA empl.			0.00969	0.00935
			(0.0182)	(0.0182)
EU empl.			0.0127	0.0140
			(0.0204)	(0.0204)
Constant	0.106***	0.137***	0.124***	0.123***
	(0.0204)	(0.0187)	(0.0233)	(0.0233)
R-squared	0.109	0.476	0.477	0.475
Observations	488	488	488	488

Note: The dependent variable in all four specifications is *MandatoryDataDisclosure*, defined as the total number of articles by authors published under a data availability policy divided by their total number of published articles. All coefficients are estimated using OLS. Standard errors are reported in parentheses.
* Significance level is reported as $p < 0.1$.
** Significance level is reported as $p < 0.05$.
*** Significance level is reported as $p < 0.01$.

resources, knowledge and experience may have an incentive to specialise in the creation of data. We argue that this specialisation may have a positive effect on the overall quality of available data.

6. Conclusion

The status quo in empirical research in economics and management is not to share data. Using a hand-collected data set consisting of 488 observations from randomly selected empirical researchers affiliated with the top 100 economics departments and top 50 business schools we show that most researchers, 394 (80.74%), do not share their data. In contrast, 82 empirical researchers (16.8%) sporadically share data, while only 12 (2.46%) share data regularly in a comprehensive and transparent way. Starting from this low data-sharing equilibrium, we empirically analyse the factors driving a researcher's choice to voluntarily share her data. First, we derive a set of testable hypotheses based on the theoretical literature on information-sharing that relates sharing behaviour to personal incentives, institutional factors and personal attitudes towards open science principles. Second, using an ordered probit model, we relate the degree of voluntary data-sharing to a set of observable characteristics of researchers. We find strong empirical support for our hypotheses that voluntary data-sharing increases with: (a) academic tenure, (b) the quality of researchers, (c) the share of published papers subject to a mandatory data-disclosure policy of journals, and (d) personal attitudes towards open science.

Despite the fact that a robust system for validating, replicating and expanding existing empirical research is in the best interest of the research community, individual researchers have insufficient incentives to provide data. The patent, cost-benefit view of data-sharing brought forward by Moffitt (2007) has been the dominant way of thinking in the empirical research community. However, the case for more data-sharing has a number of strong arguments. First, it eases replication of results. Second, it raises the credibility of empirical work – among other researchers and the public at large. Third, it increases the accuracy of research and strengthens incentives to avoid data-related errors. Fourth, it discourages fraud – not perfectly, but to a large degree. Finally, it enables researchers to use existing data sets to pursue new ideas which the originator of the data set may not have been aware of.

As researchers demand data-sharing but only very few researchers that have open science attitudes are willing to provide it, the choice to share or not to share should not be left to the discretion of individual researchers, but rather needs institutional response (Dasgupta and David, 1994). To ensure and spur progress in data-sharing, journals and universities should continue to take the lead.

Acknowledgments

We thank Badi Baltagi, Daniel Hamermesh, Dietmar Harhoff, Fabian Herweg, Mark Schankerman, Marc Scheufen, Monika Schnitzer, Olaf Siegert, Ralf Toepfer, Sven Vlaeminck, Gert G. Wagner, Joachim Wagner, Joachim Winter and three anonymous referees for their helpful comments and suggestions. We also thank participants of the Third LERU Doctoral Summer School "Beyond Open Access: Open Education, Open Data and Open Knowledge", Universitat de Barcelona, 8–13 July 2012, and seminar participants of the Micro Workshop, LMU Munich, 2 December 2013, for their valuable comments. We are very grateful to the editors of the journals under study for providing us with detailed information about the introduction (or absence) of data availability policies. We thank

Christina Wallner, Michael Gerstenberger, Benjamin Heisig, Jonas Jungbauer, Julian Hackinger, Jonas Rathfelder and Stefan Suttner for excellent research assistance. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. All data and codes we used to generate the results of this paper will be made available online for the purpose of replication or further research. We gratefully acknowledge financial support from the German Research Foundation (DFG) under the European Data Watch Extended Project. The funders had no role in study design; in the collection, analysis and interpretation of data; in the writing of the manuscript; and in the decision to submit the article for publication.

References

American Economic Review, 2013. AER Data Availability Policy, Retrieved from http://www.aeaweb.org/aer/data.php (2 April 2014).
Agarwal, R., Ohyama, A., 2013. Industry or academia, basic or applied? Career choices and earnings trajectories of scientists. Management Science 59 (4), 950–970.
Ajzen, I., 1991. The theory of planned behavior. Organizational Behavior and Human Decision Process 52 (2), 179–211.
Allan, R., 2012. Editorial: geoscience data. Geoscience Data Journal, Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/gdj.3/full (2 April 2014).
Alsheikh-Ali, A.A., Qureshi, W., Al-Mallah, M.H., Ioannidis, J.P.A., 2011. Public availability of published research data in high-impact journals. PLoS ONE 6, e24357.
Altman, M., King, G., 2007. A proposed standard for the scholarly citation of quantitative data. D-Lib Magazine 13 (3/4), Retreived from http://www.dlib.org/dlib/march07/altman/03altman.html (2 April 2014).
Anderson, R.G., Greene, W.H., McCullough, B.D., Vinod, H.D., 2008. The role of data/code archives in the future of economic research. Journal of Economic Methodology 15 (1), 99–119.
Begley, C.G., Ellis, L.M., 2012. Drug development: raise standards for preclinical cancer research. Nature 483, 531–533.
Borgmann, C.L., 2012. The conundrum of sharing research data. Journal of the American Society for Information Science and Technology 63 (6), 1059–1078.
Card, D., DellaVigna, S., 2013. Nine facts about top journals in economics. Journal of Economic Literature 51 (1), 144–161.
Christie, W.G., Schultz, P.H., 1994. Why do NASDAQ market makers avoid odd-eighth quotes? Journal of Finance 49, 1813–1840.
Cohen, W.M., Walsh, J.P., 2008. Real impediments to academic biomedical research. In: Jaffe, A.B., Lerner, J., Stern, S. (Eds.), Innovation Policy and the Economy, vol. 8. University of Chicago Press, Chicago, pp. 1–30.
Constant, D., Kiesler, S., Sproull, L., 1994. What's mine is ours, or is it? A study of attitudes about information sharing. Information Systems Research 5 (4), 400–421.
Costello, M.J., 2009. Motivating online publication of data. BioScience 59 (5), 418–427.
Coupé, T., Smeets, V., Warzynski, F., 2006. Incentives, sorting and productivity along the career: evidence from a sample of top economists. Journal of Law, Economics, and Organization 22 (1), 137–167.
Crawford, V.P., Sobel, J., 1982. Strategic information transmission. Econometrica 50 (6), 1431–1451.
Currie, G., Suhomlinova, O., 2006. The impact of institutional forces upon knowledge sharing in the UK NHS: the triumph of professional power and the inconsistency of policy. Public Administration 84 (1), 1–30.
Dasgupta, P., David, P.A., 1994. Toward a new economics of science. Research Policy 23 (5), 487–521.
Dewald, W.G., Thursby, J.G., Anderson, R.G., 1986. Replication in empirical economics: the Journal of Money, Credit and Banking project. American Economic Review 76 (4), 587–603.
DiMaggio, P.J., Powell, W.W., 1991. Introduction. In: Powell, W.W., DiMaggio, P.J. (Eds.), The New Institutionalism in Organizational Analysis. University of Chicago Press, Chicago, IL, pp. 1–38.
Economic and Social Research Council, 2010. ESRC Research Data Policy, Retrieved from http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx (2 April 2014).
European Commission, 2012. Towards Better Access to Scientific Information: Boosting the Benefits of Public Investments in Research. European Commission, Brussels.
Feigenbaum, S., Levy, D.M., 1993. The market for (ir)reproducible econometrics. Social Epistemology 7 (3), 215–232.
Ferraz, C., Finan, F., 2008. Exposing corrupt politicians: the effects of Brazil's publicly released audits on electoral outcomes. Quarterly Journal of Economics 123 (2), 703–745.
Fienberg, S.E., Martin, M.E., Straf, M.L., 1985. Sharing Research Data. National Academy Press, Washington, DC.
Furman, J.L., Stern, S., 2011. Climbing atop the shoulders of giants: the impact of institutions on cumulative research. American Economic Review 101 (5), 1933–1963.
Gans, S.J., Stern, S., 2010. Is there a market for ideas? Industrial and Corporate Change 19 (3), 805–837.
Haeussler, C., 2011. Information-sharing in academia and the industry: a comparative study. Research Policy 40 (1), 105–122.
Haeussler, C., Jiang, L., Thursby, J.G., Thursby, M., 2014. Specific and general information sharing among competing academic researchers. Research Policy 43 (3), 465–475.
Hamermesh, D.S., 1997. Some thoughts on replications and reviews. Labour Economics 4 (2), 107–109.
Hamermesh, D.S., 2006. Viewpoint: replication in economics. Canadian Journal of Economics 40 (3), 715–733.
Herndon, T., Ash, M., Pollin, R., 2013. Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. In: Political Economy Research Institute Working Paper Series, No. 322, April 2013.
Kim, Y., Stanton, J.M., 2012. Institutional and individual influences on scientists' data sharing practices. Journal of Computational Science Education 3 (1), 47–56.
Kim, Y., Stanton, J.M., 2013. Institutional and individual influences on scientists' data sharing behaviors: a multilevel analysis. In: ASIS&T 2013 Annual Meeting, Montréal, Québec, Canada, November 1–6, 2013.
Lacetera, N., Zirulia, L., 2011. The economics of scientific misconduct. Journal of Law, Economics, and Organization 27 (3), 568–603.
Levelt, Noort, Drenth Committees, 2012. Flawed Science: The Fraudulent Research Practices of Social Psychologist Diederik Stapel, Retrieved from https://www.commissielevelt.nl/ (2 April 2014).
Mann, F., von Walter, B., Hess, T., Wigand, R.T., 2009. Open access publishing in science. Communications of the ACM 52 (3), 135–139.
McCabe, M.J., Snyder, C.M., 2005. Open access and academic journal quality. American Economic Review Papers and Proceedings 95 (2), 453–458.
McCullough, B.D., 2009. Open access economics journals and the market for reproducible economic research. Economic Analysis and Policy 39 (1), 117–126.
McCullough, B.D., Vinod, H.D., 2003. Verifying the solution from a nonlinear solver: a case study. American Economic Review 93 (3), 873–892.
McCullough, B.D., McGeary, K.A., Harrison, T.D., 2006. Lessons from the JMCB Archive. Journal of Money, Credit and Banking 38 (4), 1093–1107.
McCullough, B.D., McGeary, K.A., Harrison, T.D., 2008. Do economics journal archives promote replicable research? Canadian Journal of Economics 41 (4), 1406–1420.
Merton, R.K., 1973. The sociology of science. In: Storer, N.W. (Ed.), Theoretical and Empirical Investigations. University of Chicago Press, Chicago, IL.
Mirowski, P., Sklivas, S., 1991. Why econometricians don't replicate (although they do reproduce). Review of Political Economy 3 (2), 146–163.
Moffitt, R.A., 2007. Research data integrity in economics and other social sciences. In: Presentation to National Academies Committee on Science, Engineering, and Public Policy, Committee on Assuring the Integrity of Research Data, April, 2007.
Mueller-Langer, F., Gerstenberger, M., Hackinger, J., Heisig, B., 2013. A brief guide for the creation of author-specific citation metrics and publication data using the Thomson Reuters Web of Science and Scopus databases. In: RatSWD Working Paper Series 228/2013, German Data Forum.
Mueller-Langer, F., Scheufen, M., 2013. Academic publishing and open access. In: Handke, C., Towse, R. (Eds.), Handbook on the Digital Creative Economy. Edward Elgar, Cheltenham, pp. 365–377.
Mueller-Langer, F., Watt, R., 2010. Copyright and open access for academic works. Review of Economic Research on Copyright Issues 7 (1), 45–65.
Mueller-Langer, F., Watt, R., 2014. The hybrid open access citation advantage: how many more cites is a $3,000 fee buying you? In: Max Planck Institute for Innovation & Competition Research Paper No. 14-02.
Mukherjee, A., Stern, S., 2009. Disclosure or secrecy? The dynamics of open science. International Journal of Industrial Organization 27 (3), 449–462.
National Institutes of Health, 2003. NIH Data Sharing Policy, Retrieved from http://grants1.nih.gov/grants/policy/data_sharing/ (2 April 2014).
National Science Foundation, 2011. NSF Grant Proposal Guide, Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp (2 April 2014).
Nature, 2009. Editorial, Data's shameful neglect. Nature 461 (7261), 145.
Nature, 2013. Editorial announcement: launch of an online data journal. Nature 502 (7470), 142.
Nelson, B., 2009. Empty archives. Nature 461, 160–163.
Organisation for Economic Co-operation and Development, 2007. OECD Principles and Guidelines for Access to Research Data From Public Funding. OECD, Paris.
Piwowar, H.A., Day, R.S., Fridsma, D.B., 2007. Sharing detailed research data is associated with increased citation rate. PLoS ONE 2 (3), e308.
Reinhart, C.M., Rogoff, K.S., 2013. Errata: Growth in a Time of Debt. Harvard University, Retrieved from http://scholar.harvard.edu/rogoff/publications/growth-time-debt (2 April 2014).
Rhoten, D., Powell, W.W., 2007. The frontiers of intellectual property: expanded protection vs. new models of open science. Annual Review of Law and Social Science 3, 345–373.
Roach, M., Sauermann, H., 2010. A taste for science? PhD scientists' academic orientation and self-selection into research careers in industry. Research Policy 39 (3), 422–434.
Sauermann, H., Roach, M., 2014. Not all scientists pay to be scientists: PhDs' preferences for publishing in industrial employment. Research Policy 43 (1), 32–47.
Scott, R.W., 2001. Institutions and Organizations. Sage Publications, Thousand Oaks, CA.
Siegfried, J.J., White, K.J., 1973. Financial rewards to research and teaching: a case study of academic economists. American Economic Review 63 (2), 309–315.
Stephan, P.E., 1996. The economics of science. Journal of Economic Literature 34 (3), 1199–1235.

Please cite this article in press as: Andreoli-Versbach, P., Mueller-Langer, F., Open access to data: An ideal professed but not practised. Res. Policy (2014), http://dx.doi.org/10.1016/j.respol.2014.04.008

Stern, S., 2004. Do scientists pay to be scientists? Management Science 50 (6), 835–853.
Suber, P., 2012. Open Access. MIT Press, Cambridge, MA.
Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., Frame, M., 2011. Data sharing by scientists: practices and perceptions. PLoS ONE 6 (6), 1–21.
Thursby, M., Thursby, J.G., Haeussler, C., Jiang, L., 2009. Do Academic Scientists Share Information With Their Colleagues? Not Necessarily. Vox News 29 November 2009, Retrieved from http://www.voxeu.org/article/why-don-t-academic-scientists-share-information-their-colleagues?quicktabs_tabbed_recent_articles_block=1 (2 April 2014).
Tuckman, H.P., Leahey, J., 1973. What is an article worth? Journal of Political Economy 83 (5), 951–968.
University of Pittsburgh, 2009. University of Pittsburgh Guidelines on Research Data Management. Office of the Provost, University of Pittsburgh, Pittsburgh, PA.
US House of Representatives, 2007. America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science Act, 110th Congress, 1st Session. H.R. 2272, Section 1009, Release of scientific research results. Government Printing Office, Washington.
Vlaeminck, S., Wagner, G.G., Wagner, J., Harhoff, D., Siegert, O., 2013. Replizierbare Forschung in den Wirtschaftswissenschaften erhöhen: Eine Herausforderung für wissenschaftliche Infrastrukturdienstleister. LIBREAS. Library Ideas # 23: Forschungsdaten. Metadaten. Noch mehr Daten. Forschungsdatenmanagement., pp. 29–42.
Wang, S., Noe, R.A., 2010. Knowledge sharing: a review and directions for future research. Human Resource Management Review 20 (2), 115–131.
Warlick, S.E., Vaughan, K.T.L., 2007. Factors influencing publication choice: why faculty choose open access. Biomedical Digital Libraries 4 (1), 1–12.
Weill Cornell Medical College, 2007. Research Data: Recording, Retention and Access Policy. Office of the Executive Vice Dean for Research, Weill Cornell Medical College, New York City.
Wellcome Trust, 2007. Policy on Data Management and Sharing, Retrieved from http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm (2 April 2014).
Xia, J., 2010. A longitudinal study of scholars attitudes and behaviors toward open-access journal publishing. Journal of the American Society for Information Science and Technology 61 (3), 615–624.