Vlaeminck, Sven; Wagner, Gert G.

Article

On the role of research data centres in the management of publication-related research data

LIBER Quarterly - The Journal of the Association of European Research Libraries

Suggested Citation: Vlaeminck, Sven; Wagner, Gert G. (2014) : On the role of research data centres in the management of publication-related research data, LIBER Quarterly - The Journal of the Association of European Research Libraries, ISSN 2213-056X, Igitur, Utrecht Publishing & Archiving Services, Utrecht, Vol. 23, Iss. 4, pp. 336-357, https://nbn-resolving.de/urn:nbn:NL:UI:10-1-116222

This Version is available at:
http://hdl.handle.net/11108/144

LIBER QUARTERLY
THE JOURNAL OF THE ASSOCIATION OF EUROPEAN RESEARCH LIBRARIES

Vol. 23, no. 4 (2014) 336–357 | ISSN: 1435-5205 | e-ISSN: 2213-056X

On the Role of Research Data Centres in the Management of Publication-related Research Data

Results of a survey among scientific infrastructure service providers in the field of social sciences

Sven Vlaeminck

German National Library of Economics/Leibniz Information Centre for Economics (ZBW), Hamburg
s.vlaeminck@zbw.eu

Gert G. Wagner

German Data Forum (RatSWD), German Institute for Economic Research (DIW Berlin), Max Planck Institute for Human Development, Berlin, and Berlin University of Technology (TUB)
gwagner@diw.de

Abstract

This paper summarizes the findings of an analysis of scientific infrastructure service providers (mainly from Germany but also from other European countries). These service providers are evaluated with regard to their potential services for the management of publication-related research data in the field of social sciences, especially economics. For this purpose we conducted both desk research and an online survey of 46 research data centres (RDCs), library networks and public archives; almost 48% responded to our survey. We find that almost three-quarters of all respondents generally store externally generated research data – which also applies to publication-related data. Almost 75% of all respondents also store and host the code of computation or the syntax of statistical analyses. If self-compiled software components are used to generate

Igitur publishing | http://liber.library.uu.nl/ | URN:NBN:NL:UI:10-1-116222

research outputs, only 40% of all respondents accept these software components for storing and hosting. Eight out of ten institutions also take specific action to ensure long-term data preservation. With regard to the documentation of stored and hosted research data, almost 70% of respondents claim to use the metadata schema of the Data Documentation Initiative (DDI); Dublin Core is used by 30% (multiple answers were permitted). Almost two-thirds also use persistent identifiers to facilitate citation of these datasets. Three in four also support researchers in creating metadata for their data. Application programming interfaces (APIs) for uploading or searching datasets currently are not yet implemented by any of the respondents. Least common is the use of semantic technologies like RDF.

Concluding, the paper discusses the outcome of our survey in relation to Research Data Centres (RDCs) and the roles and responsibilities of publication-related data archives for journals in the fields of social sciences.

Key Words: research data management; research data centres; journals; libraries economics

JEL Classification: C81; C88; H42; H54

1. Background and introduction

In the social sciences (especially economics, political science and sociology) more and more researchers analyse data provided by official statistics or by specialised providers of research data (e.g., from the ALLBUS at GESIS[1] or from the SOEP at DIW Berlin[2]). In addition, relevant data may often also be purchased from companies like Thomson Reuters or Bloomberg.

Especially in economics, compared to other branches of empirical research, the compilation of own datasets is not common. A major exception is the field of experimental economics, where researchers often generate their own datasets in the course of investigations motivated by game theory.

Although a rising number of publications in almost all scientific disciplines are based on the analysis of datasets, there are few effective ways to effectively replicate or re-examine the results of an empirical article, to verify it, or to make it available for re-utilisation and to support scholarly debates.

Even research data, which — in principle — is publicly available, will not typically be archived (e.g., in a final working-file) with respect to the specific selection and adjustment procedures. Therefore, while replications are not necessarily prevented, they are extremely difficult in the cases of ambitious analysis based on specific data selections and calculations.

The current situation confronts both the scientific community and scientific infrastructure service providers, like libraries and research data centres, with multiple challenges. In addition to questions concerning data availability and incentives for sharing data, there exist also infrastructure challenges. In particular the roles and responsibilities of scientific infrastructure providers, e.g., research data centres (RDCs), for managing and operating a data archive that facilitates the replications of published research often are not clearly outlined.

The first part of our paper describes some of the problems that lead to poor replicability of social sciences research. Then our paper describes the outcome of desktop research and an online survey evaluating scientific infrastructure with respect to their potential services for the management of publication-related research data in the field of social sciences.

The conclusion of our paper discusses the roles and responsibilities of several stakeholders for operating data archives for scholarly journals. Experiences in other scientific areas are integrated in our suggestions for establishing data archives that are based on the complementary know-how of research data centres (RDCs) and libraries.

1.1. Why is social science research often not replicable?

According to the literature the following reasons for missing replicability may be mentioned:

- First, and most importantly, there is a lack of incentives for researchers to share their data with the community. The academic reward system does not honour the often time-consuming efforts of data sharing — in sharp contrast to publications, although *"[a]n applied economics article is only the advertising for the data and code that produced the published results"* (Anderson, Greene, McCullough, & Vinod, 2008, p. 101).

- Furthermore, social scientists may worry that data sharing could lead to personal disadvantages. Researchers who work up and share data with the community do not receive appropriate compensation, e.g., reputation, for their efforts and may even suffer in terms of academic career because data sharing takes time that cannot be spent on research. In addition, many researchers suspect others will "misuse" their data, for example with faulty interpretations or by using a dataset without due reference to the creator of the dataset (Fecher, 2014). Eventually, the legal status of research data with regard to data sharing is not sufficiently clear, which also leads to reservations in data sharing.[3]
- Few social science journals have currently implemented guidelines requiring their authors to provide their data and statistical computation codes (McCullough, 2009; Vlaeminck, 2013). So called "data availability policies" may, in some instances, oblige the authors of empirical research papers to supply the underlying data of their results and the code/syntax of their analysis along with the manuscript of the article. Those policies often are in line with the "replication standard" formulated by Gary King (1995).
- Useful infrastructure components for the management of publication-related research data are rarely applied, which, in turn, prevents any uniform way of citing the underlying data. Available technical solutions like Dataverse,[4] a powerful tool for managing and documenting publication-related research data, are adopted by only a few journals. In this context a critical point focusses on how professional research data centres handle research-related data and what kind of services, if any, they offer.

1.2. Do research data centres offer services for archiving publication-related research data?

Research data centres could actually be ideal institutions for managing publication-related research data published as attachments to articles within scholarly journals. These capacities originate from decades of expertise in the handling of social- and economic research data, from core-competencies in the creation and maintenance of metadata collected and tagged from surveys as well as extensive experiences in managing access to data (Research Information Network, 2011). Cox and Pinfield (2013) argue that librarians,

in contrast, already often feel over-taxed with the multiple roles that they have in the various activities of their libraries. In addition, libraries may lack technical knowledge, domain-specific expertise and may also have limited personal experience in the common research processes. As such it may be difficult to position libraries as key players in this area. A loophole could be the close collaboration between libraries and research data centres to solve upcoming challenges, as Christensen-Dalsgaard (2012) suggests.

Therefore, the EDaWaX (European Data Watch Extended – www.edawax.de) project, funded by the German Research Foundation (DFG) 2011 to 2016, conducted a study evaluating if such services for publication-related research data are currently available from scientific infrastructure service providers like research data centres, libraries and archives. For this purpose a list of 46 scientific infrastructure organisations was prepared. It includes all German research data centres and data service centres accredited by the German Data Forum (RatSWD),[5] research data centres organised within the Council of European Social Science Data Archives (CESSDA),[6] the library networks in Germany as well as individual libraries and public archives.

Our investigation into the services provided by these data centres for managing publication-related research data in the social sciences is a hands-on approach to evaluate the possibility of cooperation between research libraries and research data centres. Therefore our study followed the suggestion of Lyon (2012) to develop "*a proactive approach to collaborating with disciplinary, national and international data centres … for data deposit in such archives*" (p. 130).

In a first step, the websites of these organisations were examined with regard to potential services for storing and hosting publication-related research data. The ICPRS (Inter-university Consortium for Political and Social Research — University of Michigan) provides a publication-related archive[7] that is used by numerous authors to deposit their publication-related data.[8] NARCIS,[9] a research information system located in the Netherlands, offers a specific service for publication-related data.[10] DANS EASY,[11] another service located in the Netherlands, can also be used to deposit such data in principle.[12]

However, desk research could not uncover other information needed for further analysis, which is why, in order to start a more detailed evaluation of potential services by these organisations, an online survey was conducted.

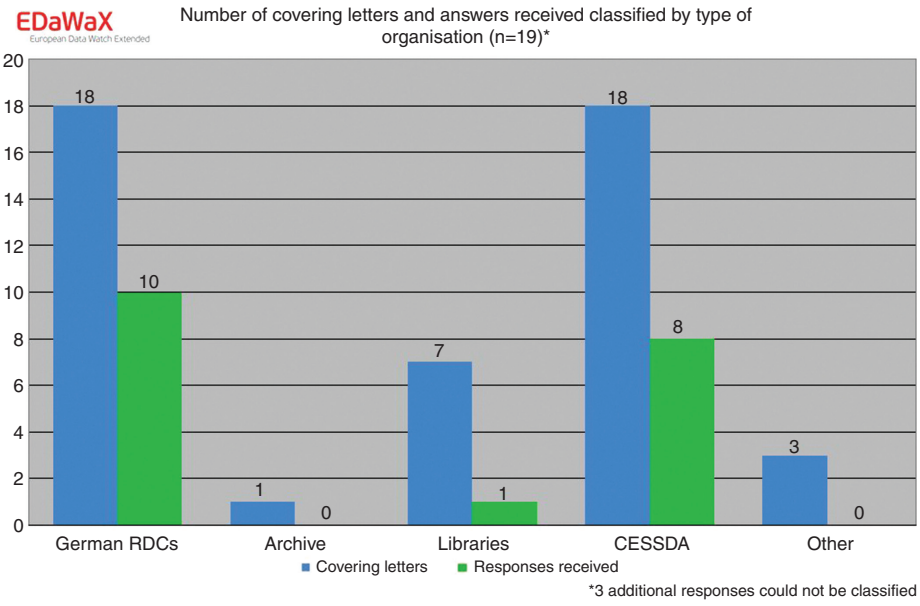
2. The online-survey

In October and November 2012 an online-questionnaire was sent to 46 organisations — among them 35 national and international research data and data service centres, 1 archive, 7 library networks and single libraries, as well as three other organisations (non-European research data centres). A satisfactory, especially when compared to average return rates of mail survey.

Due to the structure of the questionnaire, not all participating organisations responded to all questions, which explains deviations in the number of responses (Figure 1).

Certainly more important than the return rate is the structure of respondents and non-respondents. The large majority of responses came from research data centres in Germany and Europe (86%). Significantly under-represented were respondents from German library networks and archives. The three non-European research data centres did not respond.

Fig. 1: Covering letters and responses received.



We can only presume that the library networks and the archive do not offer relevant services for research data management and, therefore, did not respond to our survey.

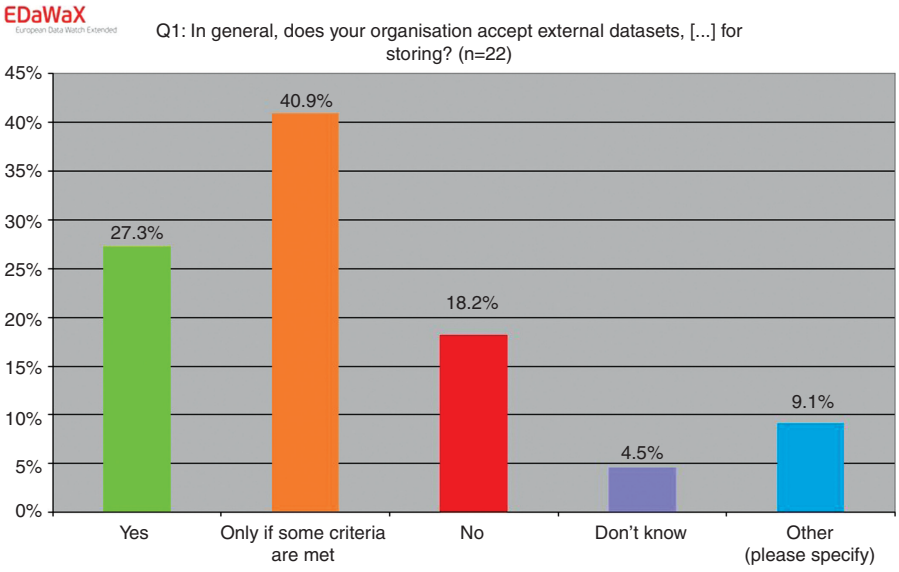
2.1. Empirical findings

Initially, the survey asked whether institutions would, in principle, host and store publication-related research data.[13] In addition, the survey also asked whether organisations would also host and store (self-compiled) software components and the code of computation/syntax of statistical analyses. These three types of data are often part of empirical submissions to economic journals.[14]

2.2. Datasets

More than three-fourths of all organisations responding accept external datasets for storage (Figure 2). At the same time the lion's share of respondents

Fig. 2: Acceptance of externally created datasets for storing.



reported that research data would only be accepted if certain criteria were met. Such criteria are subject to the specific competencies of many research data centres, but also to the specific regional/supra-regional or national competencies. Moreover, technical and organisational aspects (e.g., proper documentation, machine-readability, etc…) as well as legal concerns were cited as criteria. Approximately 74% of the respondents indicated that their organisations would host these types of data (Figure 3). If any criteria for hosting were mentioned, the subject-specific orientation of an institution was stated as main criterion.

2.3. Software

With regard to storing and hosting of (self-compiled) software components, which are often used for economic simulations, our survey indicates that just under a fourth of responding organisations accept storing and hosting software components without restrictions (Figure 4). Another 17% pointed out that they have criteria for assessing if software can be stored and hosted (e.g.,

Fig. 3: Hosting of externally created datasets.

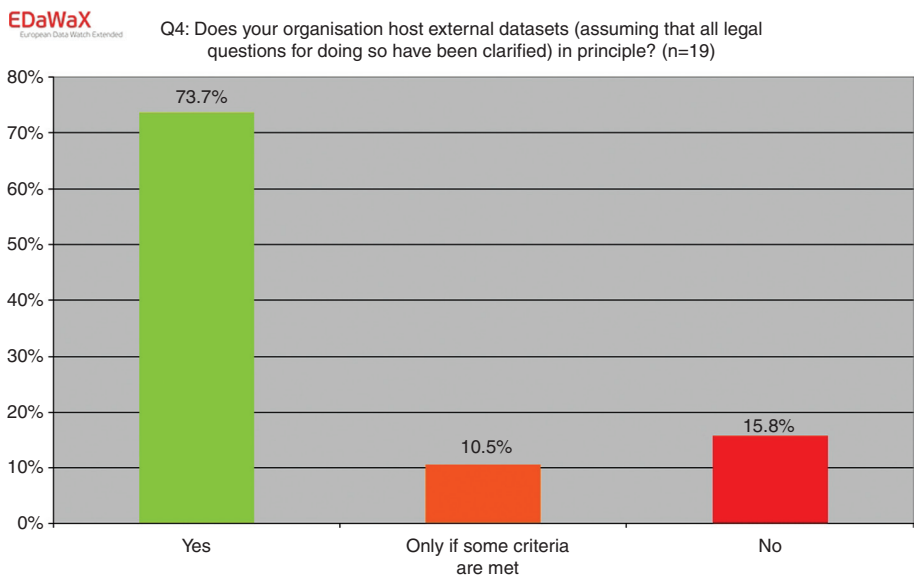
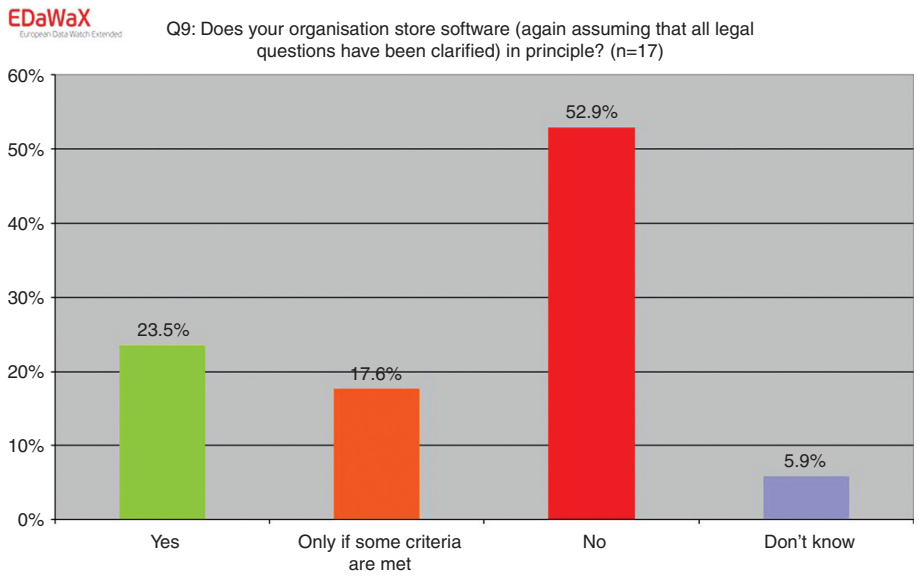


Fig. 4: Storage of software components (e.g. used for simulations).



if essential for the analysis of the data). Therefore, a gap exists in the availability of hosting and storing software components. Only a limited number of organisations offer this service.

2.4. Code of computation/Syntax of statistical analyses

Almost 70% of the organisations responding offer options to store and host computation codes (Figure 5). However, a quarter does not do so at present and is not considering offering such services in the near future. One respondent also stated a criterion — noting that the storing and hosting of these data would only be useful in the case of derived variables.

2.5. APIs

Within our analyses we also examined the availability of application programming interfaces (APIs), which enable automated data exchanges. Our results show that less than half of all responding organisations have these interfaces at their disposal (Figure 6).

Fig 5: Storage and hosting of the syntax of statistical analyses.

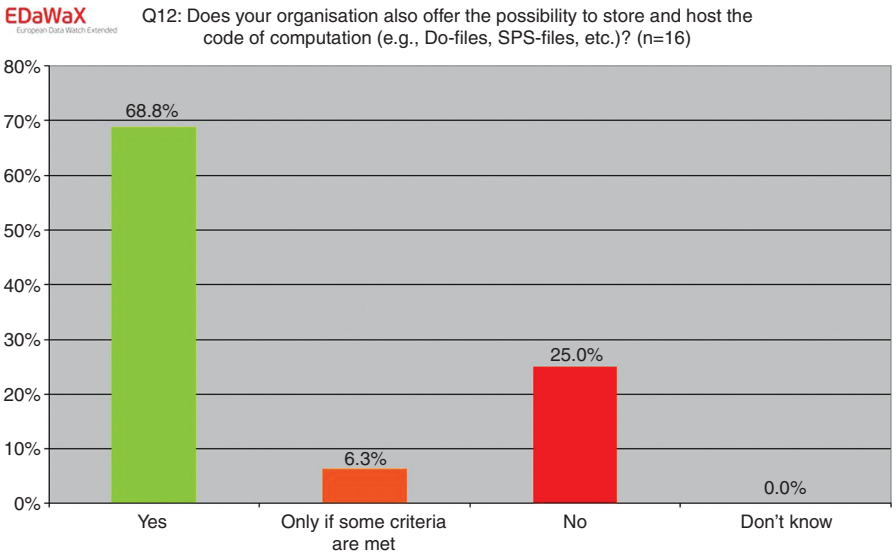
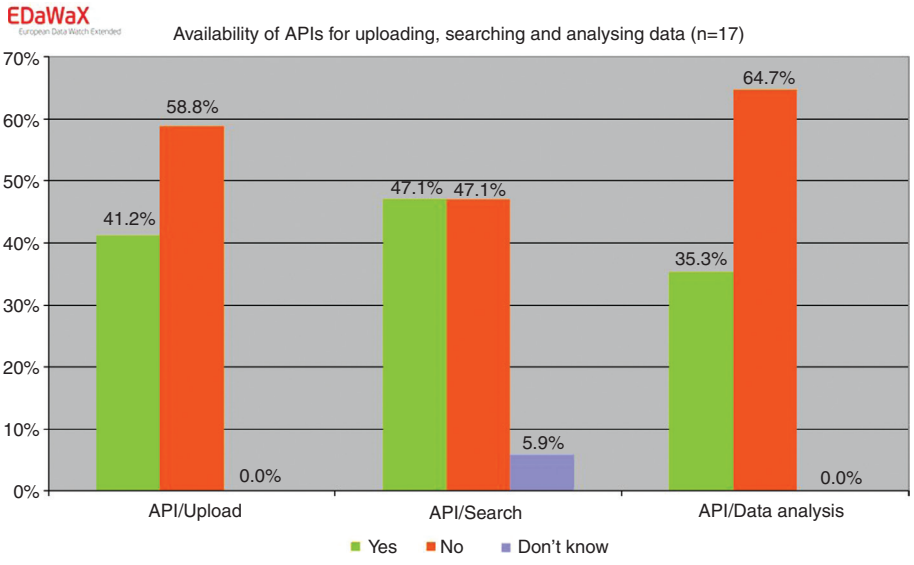


Fig. 6: Availability of APIs.



Most frequently APIs were mentioned as a device for data search (47%), followed by APIs used for uploading research data. Slightly more than a third (35%) of all respondents declared the availability of an API to analyse research data.

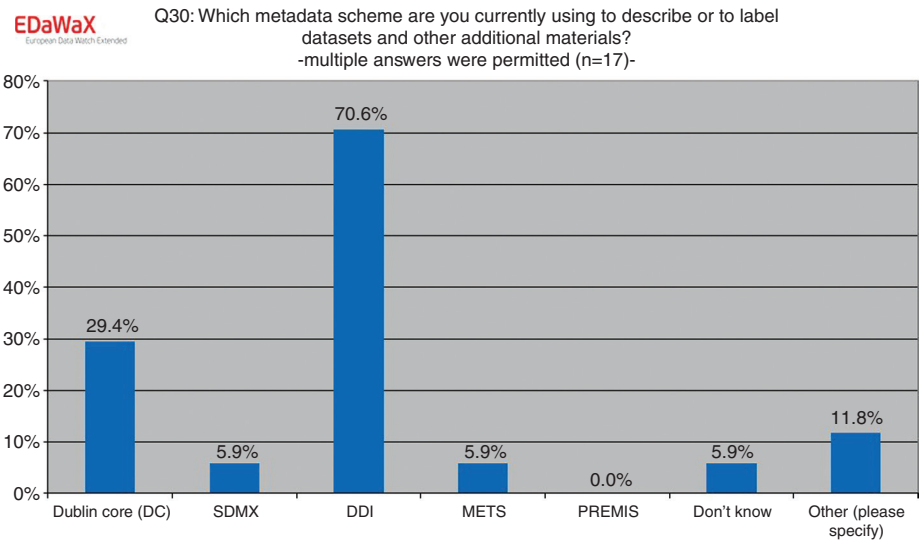
However, further analysis by EDaWaX shows that the reported interface consists only of searching and uploading interfaces on the respondents' websites. We were not able to find an API. Presumably, APIs in terms of external reading and writing accesses are by and large unknown among our respondents and not readily available.

2.6. Metadata schemata and the creation of metadata

2.6.1. Employed metadata schemata

We were also interested in the metadata schemata currently used by the organisations in their daily work. Our survey shows that more than 70% of the respondents use DDI (Figure 7). Other schemata like Dublin Core are rarely used (29%).[15] All other metadata schemata are used rather sporadically.

Fig. 7: Metadata schemata currently in use.



2.6.2. Persistent Identifiers (PI)

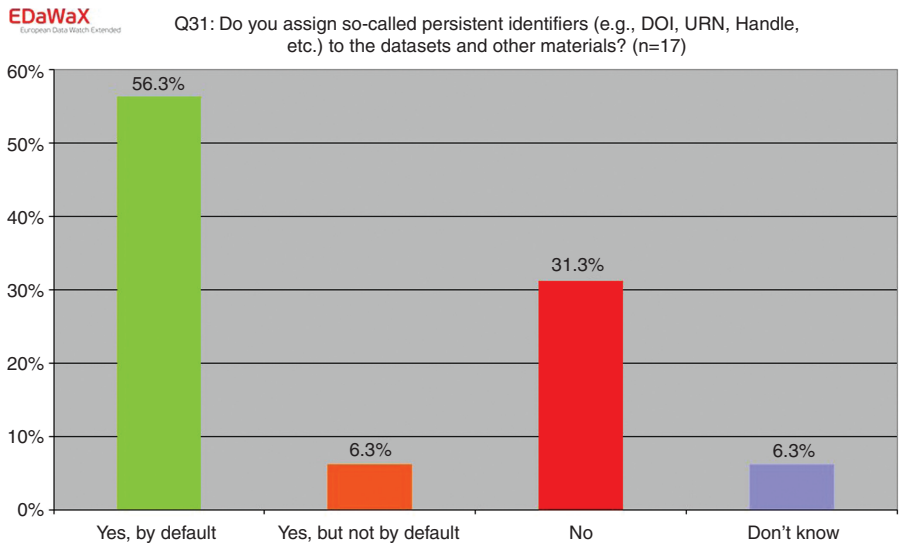
In addition, we asked, whether organisations assign persistent identifiers (e.g., handle, DOI, URN, etc…) to datasets and other materials. The persistent identification of research data is an important issue, for instance because it enables researchers to cite datasets.

More than 56% of the organisations in our sample assign such identifiers by default, but almost a third do not (Figure 8). The persistent identification of research data is an important issue, for instance because it enables researchers to cite datasets.

2.6.3. Support of Semantic Web Technologies

In our survey we also examined the implementation of RDF (Resource Description Framework). RDF is a general method for conceptual description or modelling of information implemented in web resources. Among the organisations answering this question a minority of 6% claimed to use and disseminate RDF-files. Almost a quarter of all respondents did not specify

Fig. 8: Assignment of persistent identifiers.



whether their organisation uses RDF, which presumably indicates that RDF is largely unknown.

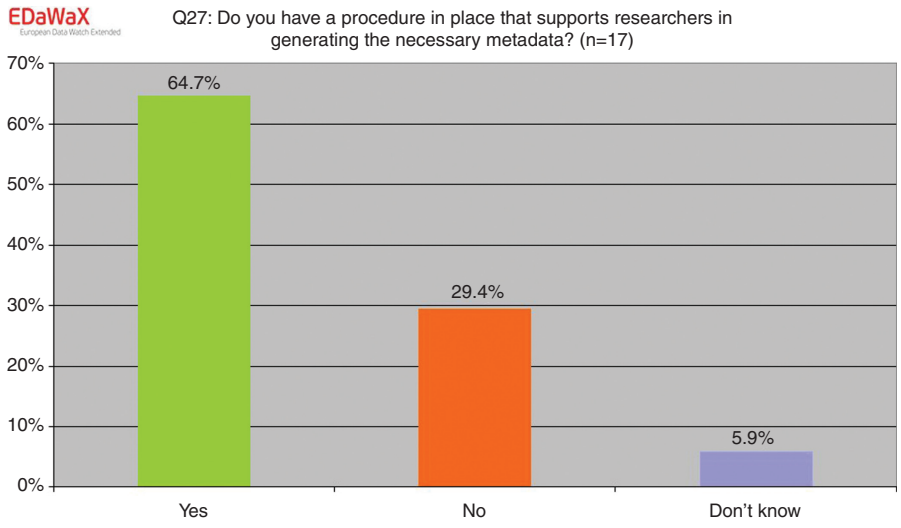
2.6.4. Support for creating metadata

Again and again, a critical issue regarding the reuse of research data is the quality of data documentation. Therefore, a matter of particular interest is whether respondents support researchers in generating metadata and, if so, how.

Our survey shows, that the majority (almost 65%) of all responding organisations do so (Figure 9).

Furthermore, we were keen to know whether this support is software-based — e.g., if there is a web frontend where researchers may type in the required information that is then converted into a standardised metadata schema.

Fig. 9: General support for metadata creation.



We find that 36% of the respondents use this type of software-based support with researchers (Figure 10).

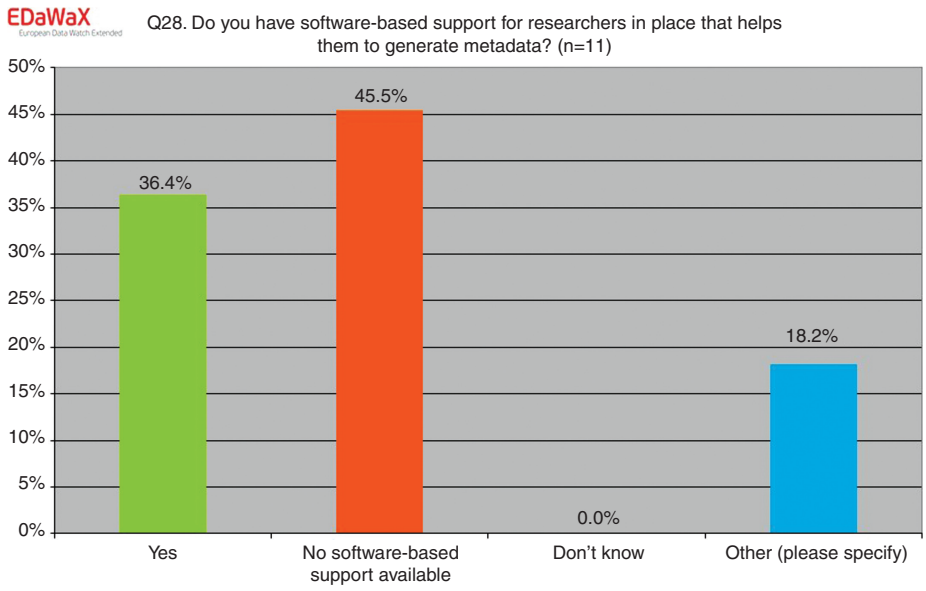
There are a striking number of statements in the section *other*. Part of the *other* support for researchers, for instance, consists of written *data deposit forms*.

Our question regarding the software program names revealed that at least two institutions use Nesstar.[16] Many organisations also use in-house solutions.

2.6.5. Digital long-term preservation

In our survey we wanted to identify to what extent the respondents' institutions have implemented specific measures for long-term research data preservation. Therefore we asked the respondents whether their organisations take specific actions for digital long-term preservation. Because format-migration is one of the dominant strategies for long-term preservation (Harvey, 2012), we suggested format migrations as one such method.

Fig. 10: Software-based support for metadata creation.



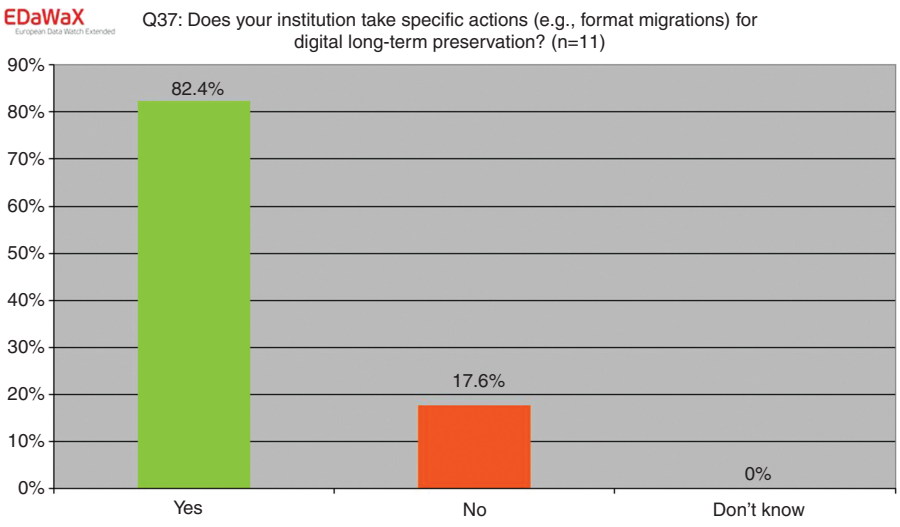
Our survey indicates that more than 80% of all organisations use these types of procedures (Figure 11).

3. Conclusion and Discussion

Our study aims to evaluate if services for publication-related research data are available from data centres, libraries and archives. Based on existing services, our project defines roles and responsibilities for operating a publication-related data archive for journals in the fields of social sciences. This approach is in line with numerous recommendations put forth by European and national organisations as well as projects to interrelate research outputs with their underlying research data (German Council of Science and Humanities, 2012; Kroes, 2012; Reilly, Schallier, Schrimpf, Smit, & Wilkinson, 2011).

A question often arising in the context of linking data and publications is the discussion about stakeholders and their roles and responsibilities in the process (Costas, Meijer, Zahedi, & Wouters, 2013; Lyon, 2007). At first glance

Fig. 11: Long-term preservation of research data.



publishers appear to be the optimal stakeholders to perform the task of building up effective and efficient data archives because many publishers already host supplementary material for their journal articles. Therefore developing and implementing data archives, collecting and disseminating research data and metadata for datasets and other material could be an easy task for publishers. Therefore wouldn't it be a good idea to rely on the publishing industry? For answering the question we have to differentiate the role of academic publishers.

On the one hand, currently publishers do not see the need to implement data archives for journals on their own (De Waard, 2012). One reason might be that implementing and operating data archives raises the costs of publication. On the other hand, the availability of a data archive does not necessarily increase the number of journal subscriptions. Hence, the incentives to build up and operate data archives are not readily apparent to publishers.

In addition, questions of ownership and access conditions to archived research data could cause uncertainty for researchers, despite a publisher's announcement "*not to require any transfer of or ownership in such data or data sets as a condition of publication of the article in question*" (STM & ALPSP, 2006, p. 1).

Despite the fact that publishers do not operate data archives for their journals, they can nevertheless play an important role in the process of interrelating research data and publications. We already observe such collaborations in some scientific disciplines where publishers and data archives actively cooperate. In disciplines like the earth sciences, the e-infrastructure needed for storing and hosting research data in conjunction with appropriate documentations of the data has already been ongoing for quite some years. From a publisher's perspective linking research data and publications provides a benefit for their journals if the scientific outputs that are enriched with research data generate more citations (Piwowar, Day, & Fridsma, 2007). In addition, these links enable a more accurate research process and offer protection against scientific misconduct (McCullough, 2009).

Excellent examples of collaborations between publishers and data repositories include PANGAEA and Dryad. PANAGEA is the "data publisher" for earth and environmental sciences. It partners with Reed-Elsevier.[17] Dryad is a non-profit repository for data underlying the international scientific and medical literature. It partners with numerous journals.[18]

Based on these experiences, the best solution is to implement and operate a discipline-specific data archive that gains importance by acquiring more and more data, which subsequently partners with publishers. The evolution and success of PANGAEA and Dryad underlines this approach impressively.

So, if it is not up to the publishers to run a data repository, other stakeholders come to the fore. In particular, research libraries and research data centres are the best positioned to take on the responsibilities of running such a disciplinary data repository. The results of our empirical investigation lead us to the conclusion that research data centres (RDCs) are likely the most relevant places to taking on the role of hosting and storing publication-related research data that is submitted to journals. RDCs already meet many prerequisites. In particular, the RDCs we analysed, in the broader field of social sciences, have much data handling experience. They are well trained in the storing, handling and documentation of these types of data as well as in taking appropriate measures for long-term data preservation.

Because RDCs currently do not comply with all requirements with respect to storing and hosting publication-related research data, collaborations between libraries and research data centres appear to be a promising way for establish such data archives: Libraries have the skills for managing publications. These include a dedicated knowledge of using authority files and multiple metadata schemata, in cataloguing information and providing this information to their discovery systems. Or as James L. Mullins, Dean of Purdue University's Library, describes it, "*Our ability to see structure to overlay on a mass of disparate 'parts,' as well as the ability to identify taxonomies to create a defined language for accessing and retrieving data is what is needed from us*" (Baykoucheva, 2011, p. 46). Unlike RDCs, it seems to be much more common for libraries to provide their stocks to their customers and to implement technical systems and the APIs necessary to do so.

According to Pullinger and Wagner (2010), managing research data comprises of a mix of information that goes beyond the traditional separate realms of publications (the primary responsibility of national libraries), official records (the responsibility of national archives) and datasets (the responsibility of researchers themselves, statistical offices and RDCs) (Pullinger & Wagner, 2010, p. 3). In addition, Cox and Pinfield (2013) emphasize that scientific libraries often do not possess specialised units experienced in both

IT-skills and knowledge in domain specific research data — a factor that hinders libraries' engagement in research data management. Establishing these departments takes time and costs money — often not attractive to scientific libraries during times of budget cuts.

Based on Lyon's suggestions to assign roles between libraries and RDCs (Lyon, 2007; adapted by Vlaeminck, 2013), we suggest the following tasks for the implementation of our project's pilot application of a data archive for economics journals, in which we strive to realise a workflow based on this division of complementary know-how.

In this distribution of tasks, ZBW — the Leibniz Information Centre for Economics — adopts the role of hosting and maintaining the metadata catalogue. Libraries then provide the technical implementation of APIs to other (library or research data) catalogues with the purpose of enriching and disseminating metadata. However one of the German RDCs, the research data centre of the Socio-Economic Panel (RDC SOEP), should take over the tasks of hosting, storing and preserving the data that has previously been submitted by editorial offices using the project's application.

By developing, implementing and operating a publication-related data archive for economics journals, both libraries and RDCs would help to ensure the validity of published economic research and to facilitate replications of these scientific outputs.

Acknowledgement

The findings presented in this article have been achieved in the course of the EDaWaX (European Data Watch Extended, www.edawax.de) research project. EDaWaX is funded by the German Research Foundation (www.dfg.de). The institutions listed hereafter are involved in the project: The German Data Forum (RatSWD), the institute Inno-tec of the LMU Munich in cooperation with the Max Planck Institute for Intellectual Property and Competition Law (IMPRS-CI) as well as the German National Library of Economics / Leibniz Information-Centre for Economics (ZBW).

In addition to the authors the following persons are involved in the EDaWaX project: Professor Klaus Tochtermann (ZBW), Professor Joachim Wagner (Leuphana

University, Lueneburg), Professor Dietmar Harhoff (IMPRS-CI and MCIER), Doctor Brigitte Preissl (ZBW), Patrick Andreoli-Versbach (IMPRS-CI), Doctor Frank Mueller-Langer (IMPRS-CI and MCIER), Olaf Siegert (ZBW), Ralf Toepfer (ZBW) and Doctor Hendrik Bunke (ZBW).

In particular the authors thank Adam Lederer (DIW Berlin) for valuable comments and suggestions for improving our English.

References

Anderson, R., 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.

Baykoucheva, S. (2011). What do libraries have to do with e-Science? An interview with James L. Mullins, Dean of Purdue University Libraries. *Chemical Information Bulletin, 63*(1), 45–49.

Christensen-Dalsgaard, B. (2012). *Ten recommendations for libraries to get started with research data management.* Final report of the LIBER working group on E-Science/ Research Data Management. Retrieved January 30, 2014, from http://www.libereurope.eu/sites/default/files/The%20research%20data%20group%202012%20v7%20final.pdf.

Costas, R., Meijer, I., Zahedi, Z., & Wouters, P. (2013). *The value of research data – Metrics for datasets from a cultural and technical point of view.* A Knowledge Exchange Report. Retrieved January 30, 2014 from www.knowledge-exchange.info/datametrics.

Cox, A.M., & Pinfield, S. (2013). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*. Published 'online before print', June 28, 2013, doi: 10.1177/0961000613492542.

De Cock Bruning, M., van Dither, B., Jeppersen de Boer, C.G., & Ringnalda, A. (2011). *The legal status of research data in the Knowledge Exchange partner countries*. Retrieved December 10, 2013, from http://www.knowledge-exchange.info/Default.aspx?ID=461.

De Waard, A. (2012). Linking data to publications: Towards the execution of papers. In: Uhlir, P. (Ed.), *For attribution – Developing data attribution and citation practices and standards: Summary of an International Workshop* (pp. 157–159). Retrieved January 30, 2014, from https://download.nap.edu/catalog.php?record_id=13564.

Fecher, B. (2014). *Data sharing angst – An insight to an ongoing research on data sharing in academia*. Alexander von Humboldt Institute for Internet

and Society. Retrieved January 30, 2014, from http://www.hiig.de/en/data-sharing-angst-an-insight-to-an-ongoing-research-on-data-sharing-in-academia/.

German Council of Science and Humanities (Wissenschaftsrat) (2012). *Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020* (No. Drs. 2359-12). Retrieved January 30, 2014, from http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf.

Guibault, L., & Wiebe, A. (2013). *Safe to be open. Study on the protection of research data and recommendations for access and usage*. Universitätsverlag Göttingen. Retrieved January 21, 2014, from http://webdoc.sub.gwdg.de/univerlag/2013/legalstudy.pdf.

Häder, M. (2009). Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland. *RatSWD Working Paper Series*, 90, Berlin. Retrieved December 10, 2013, from http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf.

Harvey, D.R. (2012). Preserving digital materials (2nd ed.). Berlin: De Gruyter.

Hillegeist, T. (2012). Rechtliche Probleme der elektronischen Langzeitarchivierung wissenschaftlicher Primärdaten, *Göttinger Schriften zur Internetforschung* (8). Retrieved December 10, 2013, from http://webdoc.sub.gwdg.de/univerlag/2012/GSI8_Hillegeist.pdf.

King, G. (1995). Replication, replication. *PS: Political Science and Politics*, 28, 443–499. Retrieved December 10, 2013, from http://gking.harvard.edu/gking/files/replication.pdf.

Kroes, N. (2012). *Commission recommendation of 17.7.2012 on access to and preservation of scientific information* [No. SWD(2012) 221 final]. Retrieved January 30, 2014, from http://dspace.utlib.ee/dspace/handle/10062/34075.

Lyon, L. (2007): *Dealing with data – Roles, rights, responsibilities and relationships*. Consultancy report. Retrieved January 30, 2014, from http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf.

Lyon, L. (2012). The informatics transform: Re-engineering libraries for the data decade. *International Journal of Digital Curation*, *7*(1). doi:10.2218/ijdc.v7i1.220.

McCullough, B.D. (2009). Open Access economics journals and the market for reproducible economic research. *Economic Analysis and Policy*, *39*(1), 117–126.

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. doi:10.1371/journal.pone.0000308.

Polhout, M. (2012). *Deposit instructions for social and behavioural sciences*. Retrieved December 10, 2013, from: http://www.dans.knaw.nl/sites/default/files/file/EASY/Deponeerinstructie%20MaGw%20UK%20DEF.pdf.

Pullinger, J., & Wagner, G. (2010). *On the respective roles of national libraries, national archives and research data centers in the preservation of and access to research data*. Retrieved January 30, 2014, from http://www.econstor.eu/dspace/handle/10419/43600.

Reilly, S., Schallier, W., Schrimpf, S., Smit, E., & Wilkinson, M. (2011). *Report on integration of data and publications.* Retrieved January 30, 2014, from http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/11/ODE-ReportOnIntegrationOfDataAndPublications-1_1.pdf.

Research Information Network (2011). *Data centres: their use, value and impact. A Research Information Network report*. Retrieved December 10, 2013, from http://www.rin.ac.uk/system/files/attachments/Data_Centres_Report.pdf.

STM & ALPSP (2006). Databases, data sets, and data accessibility – views and practices of scholarly publishers. A statement by the International Association of Scientific, Technical and Medical Publishers (STM) and the Association of Learned and Professional Society Publishers (ALPSP). Retrieved January 30, 2014, from http://www.stm-assoc.org/2006_06_01_STM_ALPSP_Data_Statement.pdf.

Vlaeminck, S. (2013). Data management in scholarly journals and possible roles for libraries – Some insights from EDaWaX. *LIBER Quarterly*, *23*(1), 48–79. URN:NBN:NL:UI:10-1-114595. Retrieved December 10, 2013, from http://liber.library.uu.nl/index.php/lq/article/view/URN%3ANBN%3ANL%3AUI%3A10-1-114595.

Notes

[1] The ALLBUS (German General Social Survey) collects up-to-date data on attitudes, behaviour, and social structure in Germany. Since 1980 representative cross-sections of the population are surveyed by GESIS every two years using both constant and variable questions. Cf. http://www.gesis.org/en/allbus/allbus-home/.

[2] The German Socio-Economic Panel Study (SOEP) is a wide-ranging representative longitudinal study of private households in Germany, in which currently about 30,000 adults living in about 15,000 households are interviewed by TNS Infratest Sozialforschung on behalf of German Institute for Economic Research (DIW) Berlin each year. Cf. http://www.soep.de.

[3] Indeed, various reports and legal opinions on research data handling have been published in recent years, but it remains questionable if the uncertainty on the part of researchers has thereby been reduced [De Cock Bruning, van Dither, Jeppersen de Boer, & Ringnalda, 2011; Guibault, & Wiebe, 2013; Häder, 2009; Hillegeist, 2012 (especially chapter A)].

[4] The website of dataverse can be found at: www.thedata.org.

[5] The website of the German Data Forum can be found at: http://www.ratswd.de/eng/index.html.

[6] The website of CESSDA can be found at: http://www.cessda.org.

[7] In the meantime, ICPSR's publication related archive changed its name to "replication datasets."

[8] A list of all journals and articles in which data stored at the ICPSR-PRA is included are available at http://www.icpsr.umich.edu/icpsrweb/ICPSR/biblio/journals?collection=DATA.

[9] The Website of NARCIS, the National Academic Research and Collaborations Information System of the Netherlands can be found at: http://www.narcis.nl/about/Language/en.

[10] More Information on NARCIS can be obtained at http://www.narcis.nl/content/pdf/narcisflyer_en.pdf. According to DANS NARCIS currently contains more than 1,800 enhanced publications and 25,000 datasets.

[11] The Website of DANS EASY can be found on: https://easy.dans.knaw.nl/ui/home.

[12] Useful information for instance was provided by Polhout (2012).

[13] Respondents had the opportunity to inspect several examples of publication-related data submissions and their elements within the online questionnaire.

[14] The required elements depend on the type of research. The data availability policy of the American Economic Review (AER) – available at http://www.aeaweb.org/aer/data.php – exemplifies such requirements.

[15] Erratum: In our survey we considered XML as a metadata schema. XML is not a metadata schema but a markup language derived from SGML. The purpose of XML is to define a set of rules for exchanging a wide variety of data. Therefore we no longer include these answers given by our respondents in our reporting.

[16] Website of nesstar, www.nesstar.com.

[17] Cf. http://www.reedelsevier.com/mediacentre/pressreleases/2010/Pages/elsevier-and-pangaea-take-next-step.aspx.

[18] Cf. http://datadryad.org/pages/integratedJournals.