Data policies in journals under scrutiny:

their strength, scope and impact

Blahous, Benedikt[1] und Gorraiz, Juan[1] und Gumpenberger, Christian[1] und Lehner, Oliver[2] und Ulrych, Ursula[1]

[1] Bibliometrics and Publication Strategies Department, Library and Archive Services, Vienna University, Vienna, Austria

[2] IT Services, Vienna University of Technology Library, Vienna University of Technology, Vienna, Austria

Abstract

Journal data policies are a potentially strong incentive for researchers to make research data available. Therefore, information about these policies is desirable. This article presents an analysis of 346 journal data policies based on a 534 cross-disciplinary journal sample with a focus on how journal publishers expect authors to make research data available. Furthermore, it includes an analysis of features such as thematic scope, user costs and hosting organisations of 171 repositories with an entry in the Thomson Reuters Data Citation Index.

DOI: 10.5283/bpf.269

URN: urn:nbn:de:bvb:355-bpf-269-6

Keywords

Research data, Journal data policies, Repositories, Data Citation Index, Citation analysis, Supplementary material

Introduction

Over the recent years, open access to research data has become a substantial issue in the scientific community.

Governments and research funding agencies all over the world have started to endorse open access policies regarding publicly funded research data [Appendix 1], and non-profit organisations, universities, research libraries and other stakeholders have launched numerous relevant projects [Appendix 2]. Journal publishers increasingly promote access to research data as well by including data policies into their author guidelines, explicitly encouraging or requesting authors to make all underlying data for their articles available. Promoting access to research data was also part of the Brussels Declaration signed by major scientific, technical and medical publishers in 2007 [1].

Open access proponents stress the potential of research data verification for the reliability of presented results. Moreover, such open data enable reuse in further projects [2-4].

Researchers recognise these advantages, but have comprehensible reservations concerning data sharing. Reasons for this include increased competition in the "publish or perish" game, legal issues such as data privacy, and the reluctance to invest additional time and effort in the preparation of data for release [5-8]. Surveys among researchers across disciplines suggest that more than 50 percent do not make their research data available [3,5,8], which is probably still a very optimistic number.

Journal data policies have been identified as a potentially strong incentive to make research data available [9,10], an assumption that is supported by several surveys among researchers [5,6,11]. However, in practice researchers are not always compliant [12,13], although making data sharing mandatory and requiring a data availability statement in the manuscript apparently boosts compliance [14].

In this context, information about the current state of journal data policies is desirable.

Studies on journal data policies go back at least to the mid-1990s, when McCain [15] analysed approximately 850 journals from sciences disciplines and identified only 132 titles, which addressed data sharing in any way. More recently, Piwowar & Chapman [16] have analysed journals with data policies on microarray data, whereas Vlaeminck & Siegert [17] focused on economics journals and Aleixandre-Benavent et al. analysed journals in the field of substance abuse [18]. Alsheikh-Ali et al. [12] and Vines et al. [14] have made studies on data policies in context with researchers' compliance based on small numbers of sciences journals. Stodden et al. [19] report on changes in data policies of 170 computational science journals. Sturges et al. [10] analysed about 400 journals from the Thomson Reuters Journal Citation Reports (JCR) 2012.

This article contributes to the available literature on journal data policies based on a cross-disciplinary journal sample (n = 534 journal titles). It provides information about which kind of data policies are common among journals (in 2014) and reports on differences observed at a high disciplinary level (sciences vs. social sciences, arts and humanities). It also deals with the question whether journal publishers prefer to manage research data by themselves or rather have it deposited in an external repository.

Furthermore, the article presents an analysis of features such as thematic scope, user costs and hosting organisations of 171 repositories indexed in the Thomson Reuters Data Citation Index (DCI). The coverage of DCI is compared with the repositories mentioned in the data policies of the journal sample. DCI, launched in 2012 as part of the Web of Science (WoS), provides an index of research data (data sets, data studies and data repositories) from across disciplines and around the world. The selection criteria are mainly based on the reputation and characteristics of the repositories [20,21]. First coverage and citation analyses of DCI have been performed from April to June 2013 by the EC3 bibliometrics group of Granada [22,23] , followed by an analysis of data citation practices [24].

Overall, the following three questions are addressed in this article:

1. Which data policies are common among journals?
2. What kind of repository is indexed in DCI?
3. Where do journal publishers recommend authors should deposit their research data and to what extent are these recommended repositories indexed in DCI?

Materials and Methods

1. Which data policies are common among journals?

The journal sample aimed at including a maximum number of journals with a data policy and at covering disciplines from the sciences, the social sciences and the arts and humanities. It was decided to combine several sources to compose the journal sample (Table 1). Thus the preselection of journals in one data pool would be attenuated and a source could be replaced with another if it turned out to be ineffective for journals from a specific discipline. The following three different types of sources were used:

- SCImago Journal & Country Rank 2012,
- repositories,
- and websites of scientific societies.

DOI: 10.5283/bpf.269

Table 1. Complete list of the sources used for composition of the journal sample

SJR 2012 (Scopus Subject Categories)	Archaeology Astronomy and Astrophysics Computer Vision and Pattern Recognition Earth and Planetary Sciences Economics Geology Law Linguistics Music Pharmaceutical Science Sociology and Political Science
Repositories	Crystallography Open Database Dryad International Food Policy Research Institute Datasets National Archive of Criminal Justice Data
Scientific societies	American Chemical Society American Economic Association American Psychological Association Ecological Society of America Public Library of Science Royal Society of Chemistry Society for American Archaeology

We used SCImago Journal & Country Rank (SJR) 2012 mainly as a source for social sciences as well as arts and humanities journals. SJR is a data portal that offers, besides other features, journal rankings according to a citation based indicator, the SJR indicator [25]. SJR is based on the coverage of the Elsevier Scopus database [26]. The top 20 to 40 journals according to the SJR indicator from 11 Scopus Subject Categories were surveyed, depending on the field. The motivation for this approach was based on the assumption that top-ranked journals in citation indexes are more likely to have a data policy, which seems to be true at least for the Journal Impact Factor [10,16,17,19]. Impact Factor rankings in WoS would have provided a comparable source. However, top journals according to the Impact Factor have been repeatedly surveyed for data policies (e.g. [12,14,18,19]), including the top 100 journals of JCR 2012 (the latest edition available at the time of the analyses for this article) by Sturges et al. [10]. Indexing over 21,000 journal titles, SCImago also has a much broader coverage than WoS with its 12,000 to 13,000 journal titles [27,28].

Repositories were identified by means of the Registry of Research Data Repositories and preselected according to their description. A repository was used if it mainly stored data related to research articles and additionally provided references to the corresponding journals. Such references were either part of the metadata of deposited data sets or available as a list on the repository website. If journal references were part of the metadata, a list of data sets was generated by performing preferably broad searches with general terms in the repository (e.g. ‘structure’ or ‘cell’). The journals in the references were then surveyed. In case that a list of journals was available, this was used as a source. In every suitable repository only a limited number of journals was surveyed, usually up to 40 like with SJR 2012.

The website of a scientific society was explored whenever an already surveyed journal published by this society had a strong or weak data policy. In this case it was likely to identify further journals of the same society with data policies. Scientific societies were also actively researched online for disciplines with low percentages of journals with a data policy in

DOI: 10.5283/bpf.269

SJR 2012 (e.g. archaeology). In general, all journals of a suitable scientific society were surveyed.

For details on the effectiveness of these search strategies see [29].

Journals which had already been surveyed by Sturges et al. [10] as part of the JoRD project were not surveyed to avoid unintended duplication. The journal list of the JoRD project has been made accessible to the authors of this article via private communication. Journals without an available website were also omitted.

For each journal surveyed in our study, the following data were collected:

- basic bibliographical information (title, ISSN etc.) and journal website URL,
- journal's main discipline according to its website and, based on that categorisation, a consecutive assignment to sciences, social sciences or arts and humanities,
- availability of a data policy,
- and, if applicable, the categorisation of the data policy.

A data policy was understood to be any statement or instruction in the author guidelines of the journal that addressed providing additional data. Instructions concerning research data (research data policies) were assigned to one of two categories, namely 'strong' and 'weak', which have been used before in similar studies [10,15,16]. Assignment was based on how closely publication was linked to data sharing. A research data policy was categorised as 'strong', if data sharing was mentioned as a requirement for publication. It was categorised as 'weak', if data sharing was merely pointed out as desideratum. In addition to research data policies, also supplementary material policies were registered. A data policy was categorised 'supplementary material' if it stated that additional materials to the article were welcome, but did not explicitly mention research data. Summing up, each identified data policy was assigned to one out of the following three categories:

- strong data policies,
- weak data policies,
- supplementary material policies.

Each journal was assigned to only one of these categories. If different statements in the author guidelines of a journal could be assigned to more than one category, the strictest was chosen (strong over weak and supplementary material, weak over supplementary material). It was decided to register supplementary material policies, because the fact that a journal facilitated data sharing, although not explicitly demanding the deposition of research data, was considered an indicator for openness towards data sharing – as also Stodden et al. [19] point out – and thus especially valuable for future analyses with focus on changes in data policies.

Data for the study were collected from the journal websites, the online periodical directory Ulrichsweb and data policies. This part of the analysis was performed in May 2014.

2. What kind of repository is indexed in DCI?

For each available repository with an entry in DCI (171 repositories then) the following data were collected:

- short description of its aims,
- assigned WoS categories,
- repository URL,
- thematic scope: Repositories were categorised as either 'disciplinary' if they accepted only data from one discipline, or as 'multidisciplinary' if they accepted data from several disciplines.
- user costs: Repositories were categorised either 'free data access, no explicit restriction of reuse' if no restrictions to data access and data reuse were mentioned, 'free data access, explicit restriction to non-commercial reuse' if only non-commercial (especially academic) reuse of stored data was allowed, and 'charges for data access or reuse' if data access or data reuse was at least partly liable to charges.
- hosting organisations: Repositories were categorised as either 'collaboration' if at least two organisations including at least one non-university served as hosting organisations, 'company' if it was a profit-oriented organisation, 'research institution' if it was a non-university research institution or a non-profit organisation and 'university'

if the repository was hosted by one or several universities.

- funding: Repositories were categorised as either 'national funding' if it got funding on national level, 'international funding' if that was the case on international level and 'funding by a company' if the repository got additional funding from a profit-oriented organisation.
- journal repositories, institutional repositories and national repositories were additionally registered.

Data were collected from the repository website, the website of the hosting organisations and, if necessary, from other online sources such as Wikipedia. This analysis was conducted in July 2014.

3. Where do journal publishers recommend authors should deposit research data and to what extent are these recommended repositories indexed in DCI?

For each journal in our sample with research data policy assignment, the data policy was revisited. Previously collected data, especially the categorisation, was rechecked. In addition, the following information was gathered:

- data policy level: Does the data policy apply for a group of the publishers' journals (e.g. of the same discipline) or just for the journal in question?
- instructions on how data should be made available: Do journal publishers prefer to manage research data themselves or have it rather deposited in an (external) repository?
- all repositories mentioned in the data policies were registered. If a repository belonged to a collaboration of repositories, this collaboration was registered instead of the repository (e.g. 'PubChem repositories' instead of 'PubChem BioAssay', 'PubChem Compound' and 'PubChem Substance'). This was decided for the purpose of standardisation. Data policies tended to name these collaborations rather than the members, although practice was inconsistent.
- for each of the repositories mentioned in the data policies, DCI was checked for a corresponding entry. This was done on repository level, considering all members of a collaboration of repositories.

This part of the analysis was conducted in December 2014. The presented results in this article are based on the revision, which agrees with the primary categorisation except for one weak data policy, which was changed into supplementary material.

Results and Discussion

1. Which data policies are common among journals?

In total the websites of 534 journals were checked for data policies. The sample included almost exclusively journals in English language from 112 publishers and 17 disciplines. The sciences and the social sciences, the arts and humanities (SAH) contributed to the sample in similar shares: 280 (52%) sciences and 254 (48%) SAH journals were analysed. The sciences section included journals from the applied sciences, astronomy, biology, chemistry, computer science, earth sciences, medical sciences, pharmaceutical science and statistics, the SAH section archaeology, criminology, economics, law, linguistics, music, psychology, sociology and political science.

346 (65%) journals were identified with an available data policy, which were categorised as follows: 68 (20%) strong data policies, 66 (19%) weak data policies and 212 (61%) supplementary material policies (Figure 1).

DOI: 10.5283/bpf.269

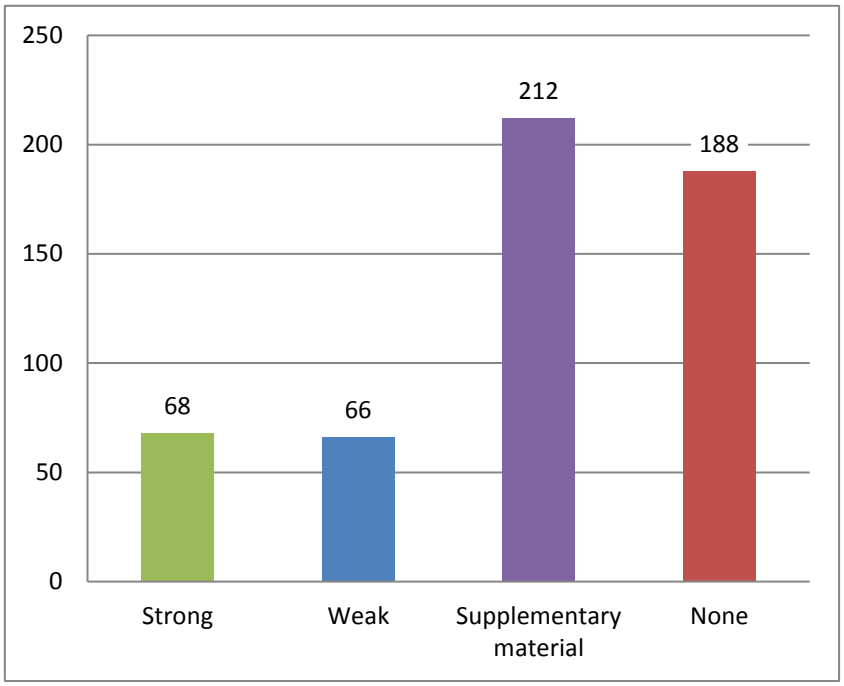

Figure 1. Categorisation of the journal sample (n = 534)

Given that our approach of data collection aimed at minimising the number of journals with no data policy in the sample, the significantly lower percentages of journals with strong or weak data policies suggest that the majority of journals do not address the availability of research data.

Sturges et al. [10] found a much higher percentage of journals with a research data policy in a cross-disciplinary sample of comparable size. About 50 percent of their 371 surveyed journals taken from the top 100 and bottom 100 from both the Science and the Social Sciences edition of JCR 2012 were categorised as having a strong or weak data policy. Assuming that journals with a top position in a citation based journal ranking are likely to have a research data policy, there might be a bias towards these journals here as well.

The relatively high percentage of supplementary material policies indicates that the majority of journals, which address data sharing in any way, do not demand it, but merely offer a non-committal service to make additional data available.

In a survey among Wiley authors conducted by the publishing group in May 2014, two thirds of the 2,250 respondents stated that they had made data available as supplementary material in a journal [5]. This suggests that supplementary material services – although obviously not designed to guarantee the availability of research data – are widely accepted and used for data sharing.

The journals in the sample with a data policy covered 51 publishers and included 192 (55%) journals from the sciences section and 154 (45%) from the SAH section. Fig 2 shows the occurrences of data policy types in both sections. In the sciences section the data policy types were relatively balanced with 57 (30%) strong data policies, 58 (30%) weak data policies and 77 (40%) supplementary material policies, whilst the SAH section had a majority of 135 (88%) supplementary material policies, followed by 11 (7%) strong and eight (5%) weak data policies.

DOI: 10.5283/bpf.269

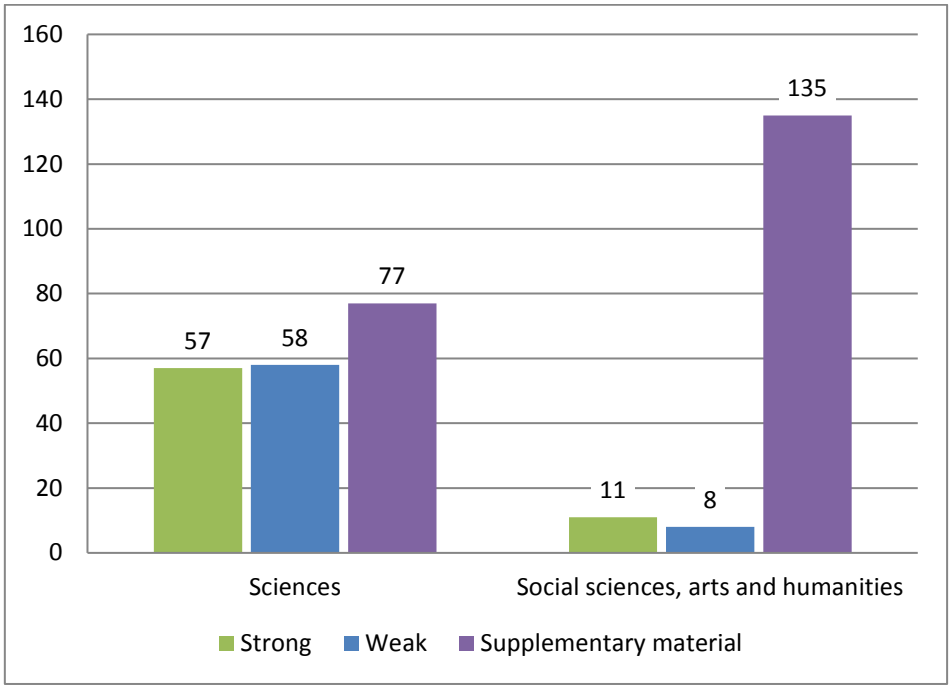

Figure 2. Occurrences of the data policy types in the two sections of the journal sample (n = 346)

Apparently, sciences journals address sharing research data considerably more frequently than journals from SAH. However, the relatively high percentage of research data policies in the sciences section did not result from equally distributed percentages in the different disciplines comprised in this section, but from peaks in the life sciences (biology, chemistry, medical sciences). In the SAH section, sociology and political science had a significantly higher percentage of research data policies than other disciplines [29].

The 134 journals in the sample with a research data policy belonged to 39 publishers. 13 (33%) of these publishers had implemented a research data policy valid either for a group or for all of their journals. Seven of these global data policies were strong, six weak (Table 2).

Table 2. Publishers represented in the journal sample with a research data policy valid for a group of journals

Strong data policy	Weak data policy
American Economic Association	Bentham Science
American Geophysical Union (Wiley-Blackwell)	BMJ Group
BioMed Central	Copernicus Publications
British Ecological Society (Wiley-Blackwell)	Ecological Society of America
Nature Publishing Group	Pensoft Publishers
Public Library of Science	Royal Society of Chemistry
Royal Society Publishing	

DOI: 10.5283/bpf.269

2. What kind of repository is indexed in DCI?

The 171 identified repositories with an entry in DCI were assigned to 84 WoS Categories. More than three quarters (77%) of these categories included only one repository, leaving 19 (23%) with at least two repositories. The three WoS Categories with the most repositories were 'Genetics & Heredity' with 39 (23%), 'Biochemistry & Molecular Biology' with 19 (11%) and 'Social Sciences, Interdisciplinary' with 16 (9%) assigned repositories. It is noteworthy that a repository can be assigned to more than one category.

More than a third of all indexed repositories were assigned to only three WoS Categories, whereas three quarters of all 84 assigned WoS Categories comprised only one repository. This suggests an inconsistent coverage of disciplines in DCI. In a former, more detailed analysis of DCI by Torres-Salinas et al. [30] in 2013, 'Genetics & Heredity', 'Biochemistry & Molecular Biology' and 'Social Sciences, Interdisciplinary' were also identified as the predominant WoS Categories, though with lower numbers of assigned repositories but in the same order. Torres-Salinas et al. (2014) also determined a bias towards disciplines from the sciences. They also found that three quarters of the other data types indexed in DCI, namely data sets and data studies, belonged to only four of the indexed repositories (Gene Expression Omnibus, UniProt Knowledgebase, PANGAEA and U.S. Census Bureau TIGER/Line Shapefiles).

In our study, 128 (75%) repositories stored data from only one discipline, whereas 43 (25%) were multidisciplinary repositories.

Of all 171 repositories, 19 (11%) served as multidisciplinary repositories on national level, two (~1%) were institutional repositories (both multidisciplinary) and one (~1%) was a disciplinary journal repository.

The hosting organisations of the repositories were in 68 (40%) cases non-university research institutions or other non-profit organisations, 55 (32%) were hosted by collaborations with at least one non-university involved, 47 (27%) by universities and one (1%) repository by a company (see Fig 3).

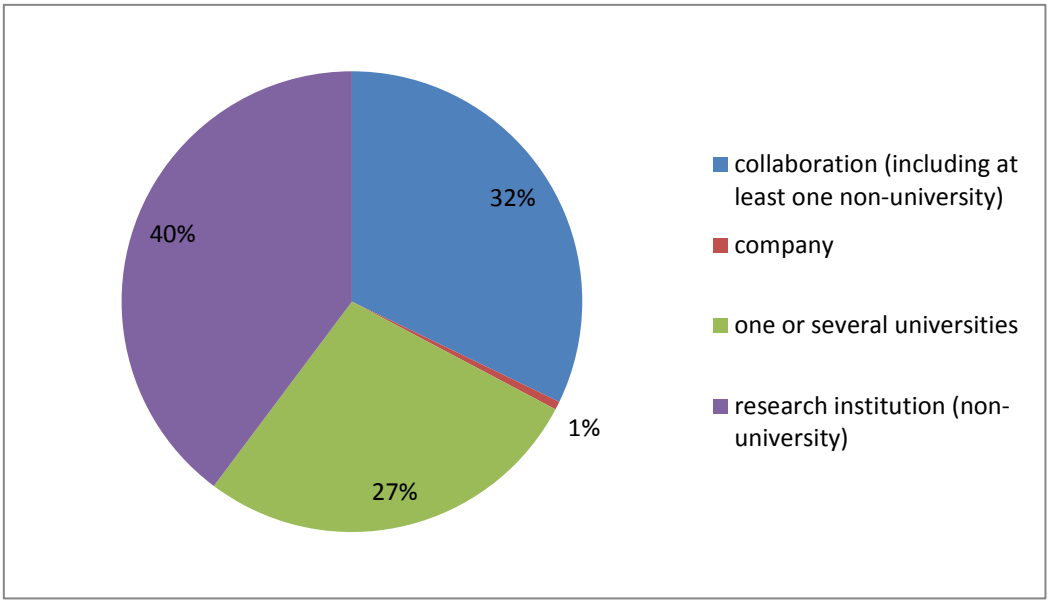

Figure 3. Distribution of hosting organisations of the repositories in DCI (n = 171)

DOI: 10.5283/bpf.269

Information about funding was not available for more than half of the repositories (60%). 50 (29%) repositories were funded on national level, 20 (12%) on international level. No repository mentioned funding by a company.

Figure 4 shows the distribution of user costs among the repositories. 167 (98%) of the repositories granted free access to stored data. 11 (7%) of these repositories explicitly restricted data reuse to non-commercial and academic purposes. Access or reuse of the data stored in 4 (2%) repositories were at least partly liable to charges.

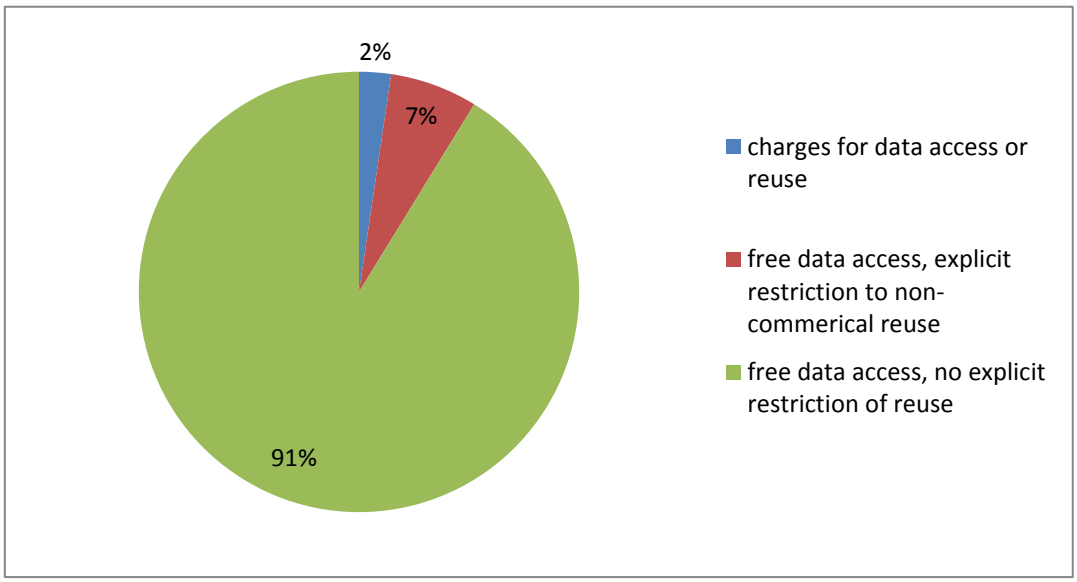

Figure 4. Distribution of user costs among the repositories in DCI (n = 171)

Summing up, these results suggest that a typical repository indexed in DCI is an open access repository, which stores data from a single discipline and is not associated with profit-oriented organisations. Journal and institutional repositories were nearly not represented. The hosting organisations were predominantly research institutions, but almost as frequently universities and collaborations. A relatively high percentage of the repositories received funding, mainly on national level.

It is not clear to what extent these results – which probably reflect some of the selection criteria of DCI – represent the repository landscape. A bias towards the sciences and an inconsistent coverage of disciplines certainly is a current feature of the repository landscape.

Repositories hosted or funded by profit-oriented organisations or publishers are probably much more frequent. The relevance of profit-oriented repositories becomes apparent by the fact that Figshare (which was not yet indexed in DCI at the time of the analysis) accounts for about a fourth of the total coverage in the database in December 2014 [31]. On the other hand, most journal repositories might be too small to meet the selection criteria of DCI concerning coverage or range of influence. The same seems to be true for institutional repositories.

3. Where do journal publishers recommend authors should deposit research data and to what extent are these recommended repositories indexed in DCI?

The distribution of the journals' instructions on how to make research data available is shown in Figure 5. Of the 134 journals in the sample with a research data policy, 105 (78%) requested the deposition of research data in an external repository. The remaining 29 (22%) journals requested that research data should be sent directly to them who would make it available via the journal website or the journal publisher's repository.

97 (72%) data policies mentioned at least one specific repository suitable for data deposition.

On average, one of these 97 data policies named nine repositories. The three repositories with the most mentions were GenBank with a total of 72, Gene Expression Omnibus with 59 and the European Nucleotide Archive with 57 mentions.

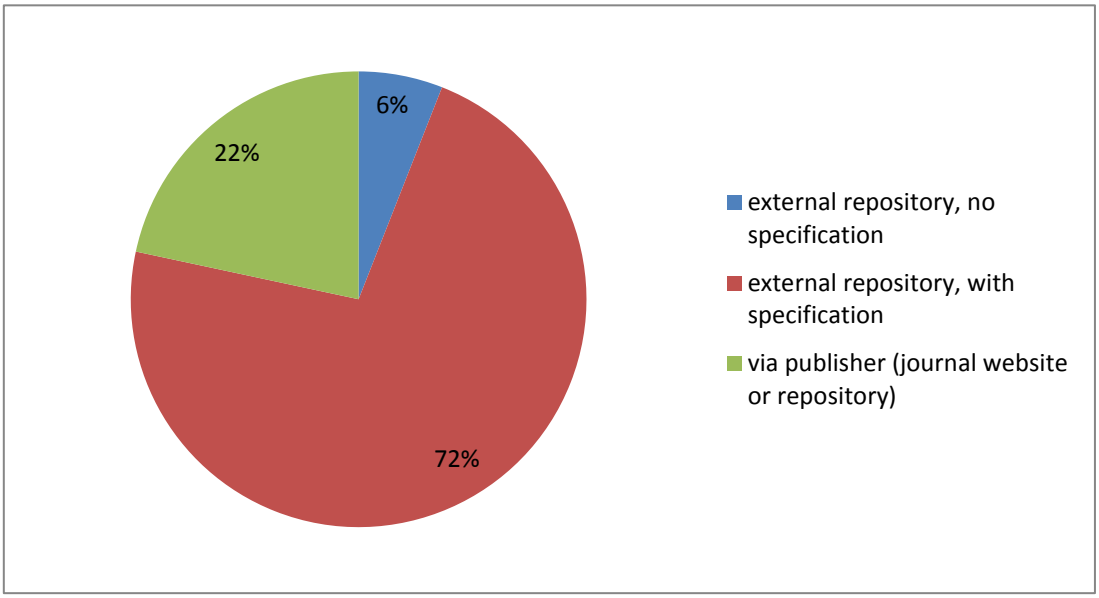

Figure 5. Distribution of journals' instructions on how to make research data available (n = 134)

21 (20%) of the repositories mentioned had an entry in DCI. Table 3 shows the total number of mentions for all repositories with five or more mentions, and whether the repository had an entry in DCI. In total, there were 839 mentions of repositories in the 97 data policies of which 408 (49%) named one of the 21 repositories indexed in DCI.

DOI: 10.5283/bpf.269

Table 3. Repositories with five or more mentions in the data policies that specified where to deposit data (n = 97)

Repository denomination	Total number of mentions	Entry in DCI
GenBank	72	No
Gene Expression Omnibus	59	Yes
European Nucleotide Archive	57	Yes
DNA Data Bank of Japan	56	No
ArrayExpress	51	Yes
Worldwide Protein Data Bank	46	Yes*
Cambridge Structural Database	44	Yes
UniProt	44	No
Nucleic Acid Database	39	Yes
PubChem	38	No
BioModels Database	36	No
ChemSpider	35	No
Dryad	33	Yes
Knowledge Network for Biocomplexity	14	Yes
TreeBASE	14	Yes
NERC Data Centres	12	No
Oak Ridge National Laboratory Distributed Active Archive Center	11	Yes
PANGEA	8	Yes
Atmospheric Radiation Measurement Data Archive	7	Yes
Carbon Dioxide Information Analysis Center	7	No
ClinicalTrials.gov	7	No
Figshare	7	Yes
Marine Geoscience Data System	7	No
EarthChem Library	6	No
GeoScenic	6	No
System for Earth Sample Registration	6	No
Woods Hole Open Access Server	6	Yes
International Tree-Ring Data Bank	5	No
Kristallstruktur-Depot (FIZ Karlsruhe)	5	No
*The Protein Data Bank and the Biological Magnetic Resonance Data Bank were indexed.		

A vast majority (92%) of the journals, which expected authors to deposit data externally, named specific repositories.

In contrast, Sturges et al. [10] found only about 15 percent of the identified data policies in their sample to name specific repositories, and 17 percent expecting the deposition in a repository but without specification. Even assuming that Sturges et al. included supplementary material policies in their strong and weak data policies, this suggests recent

changes in journal data policies towards more specific instructions on data deposition.

The repositories mentioned in the data policies of the journal sample were predominantly disciplinary repositories from the life sciences, which agrees with the coverage in DCI. Repositories from the social sciences were hardly mentioned. With 20 percent, the number of repositories represented in DCI was relatively small, which is probably due to the broader coverage of disciplines in DCI. A surprising result is that some of the repositories with the most mentions in the data policies were not indexed, among them the repository with the most mentions, GenBank, and other repositories from the life sciences like the DNA Data Bank of Japan, UniProt and the PubChem repositories. Since *Genetics & Heredity* was the category in DCI which accounted for the most repositories, any selection is obviously difficult within this discipline due to the large number of already established repositories. On the other hand, almost half (49%) of all the mentions in the data policies referred to DCI indexed repositories, which proves a database indexing policy of in fact well-established repositories.

Instructions on how to make research data available differed significantly in the sciences and the SAH section (Fig 6). In the sciences section 96 (84%) journals were categorised 'external repository, with specification', five (4%) 'external repository, no specification' and 14 (12%) 'via publisher (journal website or repository)'. In the SAH section one (5%) journal was categorised 'external repository, with specification', three (16%) 'external repository, no specification' and 15 (79%) 'via publisher (journal website or repository)'.

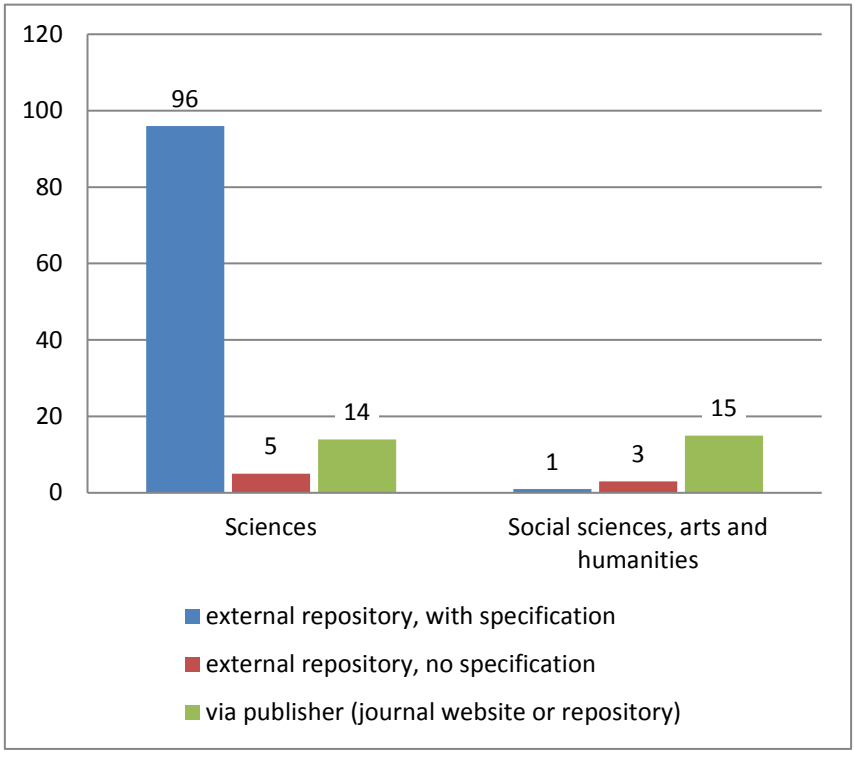

Figure 6. Comparison of both sections of the journal sample concerning journals' instructions on how to make research data available (n = 134)

These results suggest that journals from the sciences with a research data policy have a clear preference for an external deposition of research data. A likely explanation is that, especially in the life sciences, disciplinary repositories have already been established for several decades (e.g. the American Type Culture Collection or the Protein Data Bank), which certainly encourages external deposition. In addition, sciences journals would need to manage by trend large amounts of data and specialised data formats. This fact also

DOI: 10.5283/bpf.269

amounts to an increased attractiveness of external data deposition.

The number of journals from the SAH section with a research data policy in our study is certainly too small to allow generalization. Nevertheless, also including the large number of supplementary material policies in this section, apparently the conditions in SAH differ from the ones in the sciences: few well-established repositories, smaller amounts of data and more conventional data formats allow for an easier research data management, which is hence more attractive for journal publishers. Nevertheless, this interpretation seems inappropriate for some disciplines from the social sciences with a strong empirical background (e.g. sociology, political science, economics), which certainly deal with considerable amounts of data such as survey and census data [31]. However, many data generated in these disciplines might not be included in journal publications, but rather disseminated via other publication channels such as government reports. Results presented in journal publications are often based on smaller data sets (subsets of the overall generated research data), which can either be presented as supplementary materials or included in the publication as appendices.

Conclusions

Compared to the number of active journals, the number of journals considered in this analysis is relatively small due to limited time and staff resources and the large amount of data collection, which had to be done manually. Therefore, the study was not designed to provide statistically accurate results, but merely to inform about the current state of journal data policies.

Accordingly, the results indicate the following situation: Research data policies are still relatively rare, while supplementary material policies are well established. The proportion of required versus encouraged data deposition is balanced. A detailed study on which kinds of data policies ensure compliance, but also on the effect of supplementary material policies, is desirable.

In general, the results suggest recent changes towards more specific instructions in journal data policies on where to deposit research data. In comparison, journals from sciences disciplines have a significantly higher number of research data policies than journals from the social sciences or the arts and humanities. In addition, deposition of research data in an external repository is more common with journals from sciences disciplines. Due to the low number of journals from the social sciences in the sample with a research data policy, the observed results require further studies on this subject.

The analysis of the repositories with an entry in DCI suggests that the majority of indexed repositories are open access, disciplinary repositories. It also confirms a bias towards the coverage of sciences disciplines. Moreover, the number of indexed repositories per discipline varies considerably.

Only a relatively low percentage of the repositories mentioned in the research data policies of the journal sample had an entry in DCI, which is probably due to the broader coverage of disciplines in DCI. However, the repositories with an entry in DCI were mostly repositories with above-average numbers of mentions in the data policies.

There is currently no comprehensive information source for evaluation of repositories. Nevertheless, the Registry of Research Data Repositories (www.re3data.org) has by now indexed more than 1,000 repositories and encompasses for each repository features like content types, repository size, responsible institutions and legal aspects such as conditions for data access. With this information on hand, it has the potential to be a suitable basis for the evaluation and selection of repositories for citation indexes.

Acknowledgments

These analyses were done within the scope of e-infrastructures Austria (http://e-infrastructures.at/).

The authors thank Dr. Uwe Wendland (Thomson Reuters) for granted trial access to the Data Citation Index.

Dr. Marianne Bamkin and Azhar Hussain from the JoRD project (https://jordproject.wordpress.com/) provided us with data from their project via private communications, before the corresponding article was available. This was very helpful for the general framework of this analysis, and we are highly grateful for the kind cooperation.

Finally, the authors also thank Mag. Bettina Stein, who performed a considerable amount of the data collection for the journal sample.

References

Appendix 1. Hyperlinks to open access policies of governments and funding agencies

- African Development Bank Group. Open Data for Africa – Data Submission & Dissemination Tool User Manual [Internet]. Abidjan 01, Côte d'Ivoire: African Development Bank Group; c2015 [cited 8 December 2015]. Available from: http://www.afdb.org/en/knowledge/statistics/open-data-for-africa/.
- Max Planck Society. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities [Internet]. Munich, Germany: Max Planck-Society; c2003-20162003 October 22; [cited 8 December 2015] Available from: http://openaccess.mpg.de/67605/berlin_declaration_engl.pdf.
- Department of Biotechnology and Department of Science & Technology, Ministry of Science & Technology, Government of India. DBT and DST Open Access Policy – Policy on open access to DBT and DST funded research [Internet]. 2014 December 12 [cited 8 December 2015]. Available from: http://www.dst.gov.in/sites/default/files/APPROVED%20OPEN%20ACCESS%20POLICY-DBT%26DST%2812.12.2014%29_1.pdf.
- European Commission. Open data – An engine for innovation, growth and transparent governance [Internet]. Brussels, Belgium: European Commission; 2011 December 12 [cited 8 December 2015]. COM(2011) 882. Available from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF.
- Rieck K, Reckling F. Open Access Policy for FWF-funded projects [Internet]. Vienna, Austria: Austrian Science Fund (FWF); [cited 8 December 2015]. Available from: https://www.fwf.ac.at/en/research-funding/open-access-policy/.
- Iniciativa Latinoamericana de Datos Abiertos. About ILDA [Internet]. N.a.: Iniciativa Latinoamericana de Datos Abiertos; [cited 8 December 2015]. Available from: http://idatosabiertos.org/about-ilda/.
- Research Councils UK. RCUK Common Principles on Data Policy [Internet]. Swindon, UK: Research Coucils UK; 2015 July [cited 8 December 2015]. Available from: http://www.rcuk.ac.uk/research/datapolicy/.
- The Caribbean Open Institute. Open Data [Internet]. N.a.: The Caribbean Open Institute; c2012 [cited 8 December 2015]. Available from: http://www.caribbeanopeninstitute.org/content/open-data.
- Government of Japan/IT Strategic Headquarters. Open Government Data Strategy [Internet]. 2012 July 4 [cited 8 December 2015]. Available from: http://japan.kantei.go.jp/policy/it/20120704/text.pdf.
- The National Science Foundation. Proposal and Award Policies and Procedures Guide. Part II – Award & Administration Guide [Internet]. NSF 15-1, effective December 26, 2014 [cited 8 December 2015]. Available from: http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/aag_print.pdf.

Appendix 2. Hyperlinks to projects related to research data

- recodeproject.eu [Internet]. London, UK: Trilateral Research & Consulting, LLP; c2013 [cited 8 December 2015]. Available from: http://recodeproject.eu/.
- alliancepermanentaccess.org [Internet]. The Hague, The Netherlands: Alliance for Permanent Access; [cited 8 December 2015]. Available from: http://www.alliancepermanentaccess.org/.
- codata.org [Internet]. Paris, France: CODATA; c2015 [cited 8 December 2015]. Available from: http://www.codata.org/.
- re3data.org [Internet]. Karlsruhe, Germany: Karlsruhe Institute of Technology; [cited 8 December 2015]. Available from: http://www.re3data.org/.
- okfn.org [Internet]. Cambridge, UK: St John's Innovation Centre; [cited 8 December 2015]. Available from: https://okfn.org/.
- rd-alliance.org [Internet]. N.a.: Research Data Alliance; [cited 8 December 2015]. Available from: https://rd-alliance.org/.
- datacite.org [Internet]. Hannover, Germany: German National Library of Science and Technology; [cited 8 December 2015]. Available from: https://www.datacite.org/.

Literature

1. International Association of Scientific, Technical and Medical Publishers. Brussels Declaration on STM Publishing. Position Paper [Internet]. 2007 November 1 [cited 8 December 2015]. Available from: http://www.stm-assoc.org/2007_11_01_Brussels_Declaration.pdf.
2. Borgman CL. Research Data: Who will share what, with whom, when, and why? China-North America Library Conference; 2010 September 8-12; Beijing, PRC. Available from: http://works.bepress.com/borgman/238.
3. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read L et al. Data Sharing by Scientists: Practices and Perceptions. PLoS ONE. 2011;6(6): e21101. doi: 10.1371/journal.pone.0021101.
4. Uhlir PF, Schröder P. Open Data for Global Science. Data Sci J. 2007 June 17;6:OD36-OD53. doi: http://dx.doi.org/10.2481/dsj.6.OD36.
5. Ferguson L. How and why researchers share data (and why they don't). 2014 Nov 3 [cited 8 December 2015] In: Wiley Exchanges Blog [Internet]. Available from: http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/.
6. Kim Y, Stanton JM. Institutional and Individual Influences on Scientists' Data Sharing Practices. Journal of Computational Science Education, 2012 June;3(1):47-56.
7. Pienta AM, Alter G, Lyle G. The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data. The Organisation, Economics and Policy of Scientific Research Workshop; 2010 April; Torino, Italy. Available from: http://deepblue.lib.umich.edu/handle/2027.42/78307.
8. Kuipers T, Van der Hoeven J. PARSE: Insight into digital preservation of research output in Europe. Survey report. Brussels, Belgium: European Commission. 2009 December. Project: FP7-2007-223758 PARSE.Insight. Deliverable number D3.4. Available: http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf.
9. Lin J, Strasser C. Recommendations for the Role of Publishers in Access to Data. PLoS

   Biol. 2014;12(10): e1001975. doi: 10.1371/journal.pbio.1001975.
10. Sturges P, Bamkin M, Anders JHS, Hubbard B, Hussain A, Heeley M. Research Data Sharing: Developing a Stakeholder-Driven Model for Journal Policies. J Am Soc Inf Sci Technol. 2015 April 7;66(12):2445-2455. doi: 10.1002/asi.23336.
11. Pham-Kanter G, Zinner DE, Campbell EG. Codifying Collegiality: Recent Developments in Data Sharing Policy in the Life Sciences. PLoS ONE. 2014;9(9): e108451. doi: 10.1371/journal.pone.0108451.
12. Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JPA. Public Availability of Published Research Data in High-Impact Journals. PLoS ONE. 2011;6(9): e24357. doi: 10.1371/journal.pone.0024357.
13. Savage CJ, Vickers AJ. Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. PLoS ONE. 2009;4(9): e7078. doi: 10.1371/journal.pone.0007078.
14. Vines TH, Andrew RL, Bock DG, Franklin MT, Gilbert KJ, Kane NC et al. Mandated data archiving greatly improves access to research data. FASEB J. 2013;27:1304-1308. doi: 10.1096/fj.12-218164.
15. McCain KW. Mandating Sharing. Journal Policies in the Natural Sciences. Science Communication. 1995 June;16(4):403-431.
16. Piwowar HA, Chapman WW. A review of journal policies for sharing research data. In: Chan L, Mornati S, editors. ELPUB2008. Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0 - Proceedings of the 12th International Conference on Electronic Publishing; 2008 June 25-27; Toronto, Canada. Available from: http://elpub.scix.net/cgi-bin/works/Show?001_elpub2008.
17. Vlaeminck S, Siegert O. Welche Rolle spielen Forschungsdaten eigentlich für Fachzeitschriften? Eine Analyse mit Fokus auf die Wirtschaftswissenschaften. Working Paper. Federal Ministry of Education and Research (Germany). 2012 November. RatSWD Working Paper Series 210/2012. German. Available from: http://www.ratswd.de/publikationen/working-papers/2012.

18. Aleixandre-Benavent R, Vidal-Infer A, Alonso Arroyo A, Valderrama Zurián JC, Bueno Cañigral F, Ferrer Sapena A. Public availability of published research data in substance abuse journals. International Journal of Drug Policy. 2014 November;25(6):1143-1146. doi: 10.1016/j.drugpo.2014.07.007

19. Stodden V, Guo P, Ma Z. Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals. PloS ONE. 2013;8(6): e67111. doi: 10.1371/journal.pone.0067111.

20. Breen J, Anderson A, Thomson Reuters. CORRECTION: Thomson Reuters Launches Data Citation Index for Discovering Global Data Sets. Press release [Internet]. 2012 October 17 [cited 8 December 2015]. Available from: http://www.reuters.com/article/2012/10/17/idUS187519+17-Oct-2012+HUG20121017.

21. Thomson Reuters. Repository evaluation, selection and coverage policy for the Data Citation Index [Internet]. 2012 [cited 8 December 2015]. Available from: http://wokinfo.com/media/pdf/DCI_selection_essay.pdf.

22. Torres-Salinas D, Jiménez-Contreras E, Robinson-García N. How many citations are there in the Data Citation Index? Proceedings of the science and technology indicators conference; 2014 September 3-5; Leiden, The Netherlands. Leiden: Universiteit Leiden; 2014.

23. Torres-Salinas D, Robinson-García N, Cabezas-Clavijo Á (2013) Compatir los datos de investigación: Una introducción al 'Data Sharing'. El profesional de la información. 2012 March-April;21(2):173-184. Spanish.

24. Robinson-García N, Jiménez-Contreras E, Torres-Salinas D. Analyzing data citation practices using the data citation index. Journal of the Association for Information Science and Technology. 2016 December; 67(12): 2964-2975. doi: 10.1002/asi.23529

25. González-Pereira B, Guerrero-Bote VP, Moya-Anegón F (2010) A new approach to the metric of journals' scientific prestige: The SJR indicator. Journal of Infometrics. 2010 July;4(3):379-391. doi: 10.1016/j.joi.2010.03.002.

26. scimagojr.com [Internet]. About us. SCImago Journal & Country Rank. 2007 [cited 8 December 2015]. Available from: http://www.scimagojr.com/aboutus.php.

27. Scopus. Content Coverage Guide [Internet]. 2016 January. [cited 15 April 2016]. Available from: https://www.elsevier.com/__data/assets/pdf_file/0007/69451/scopus_content_coverage_guide.pdf.

28. Thomson Reuters (2012) Web of Science. The definitive resource for global research [Internet]. 2012 [cited 8 December 2015]. Available from: http://wokinfo.com/media/pdf/WoSFS_08_7050.pdf.

29. Blahous B, Gorraiz J, Gumpenberger C, Lehner O, Stein B, Ulrych U. Forschungsdatenpolicies in wissenschaftlichen Zeitschriften – Eine empirische Untersuchung. Zeitschrift für Bibliothekswesen und Bibliographie. 2015 February 13;62(1):12-24. German.

30. Torres-Salinas D, Martín-Martín A, Fuente-Gutiérrez E. Analysis of the coverage of the Data Citation Index – Thomson Reuters: disciplines, document types and repositories. Revista Española de Documentación Científica. 2014;37(1): e036. doi: http://dx.doi.org/10.3989/redc.2014.1.1114.

31. Peters I, Kraker P, Lex E, Gumpenberger C, Gorraiz J. Research Data Explored: Citations versus Altmetrics. Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Infometrics Conference; 2015 June 20 to July 3; Istanbul, Turkey. Istanbul: Bogaziçi University Printhouse; 2015.