RESEARCH PAPER

Recommended versus Certified Repositories: Mind the Gap

Sean Edward Husen[1], Zoë G. de Wilde[1], Anita de Waard[2] and Helena Cousijn[3]

[1] Leiden University, Rapenburg 70, 2311 EZ Leiden, NL

[2] Elsevier, 71 Hanley Lane, Jericho, VT 05466, US

[3] Elsevier, Radarweg 29, 1043NX, Amsterdam, NL

Corresponding author: Sean Edward Husen (s.e.husen@hum.leidenuniv.nl)

Researchers are increasingly required to make research data publicly available in data repositories. Although several organisations propose criteria to recommend and evaluate the quality of data repositories, there is no consensus of what constitutes a good data repository. In this paper, we investigate, first, which data repositories are recommended by various stakeholders (publishers, funders, and community organizations) and second, which repositories are certified by a number of organisations. We then compare these two lists of repositories, and the criteria for recommendation and certification. We find that criteria used by organisations recommending and certifying repositories are similar, although the certification criteria are generally more detailed. We distil the lists of criteria into seven main categories: "Mission", "Community/ Recognition", "Legal and Contractual Compliance", "Access/Accessibility", "Technical Structure/ Interface", "Retrievability" and "Preservation". Although the criteria are similar, the lists of repositories that are recommended by the various agencies are very different. Out of all of the recommended repositories, less than 6% obtained certification. As certification is becoming more important, steps should be taken to decrease this gap between recommended and certified repositories, and ensure that certification standards become applicable, and applied, to the repositories which researchers are currently using.

Keywords: data repositories; data management; certification; data quality; data fitness

Introduction

Data sharing and data management are topics that are becoming increasingly important. More information is appearing about their benefits, such as increased citation rates for research papers with associated shared datasets (Piwowar & Vision 2013; Piwowar, Day & Fridsma 2007). A growing number of funding bodies such as the NIH and the Wellcome Trust (NIH 2015; Wellcome Trust n.d. a), but also several journals (Borgman 2012), have installed policies that require research data to be shared (Mayernik et al. 2014). To be able to share data, both now and in the future, datasets not only need to be preserved, but also need to be comprehensible, and useable for others. To ensure these qualities, research data needs to be managed (Dobratz et al. 2010) and data repositories can play a role in maintaining the data in a useable structure (Assante et al. 2016). However, using a data repository does not guarantee that the data is usable, since not every repository uses the same procedures and quality metrics, such as applying proper metadata tags (Merson, Gaye & Guerin 2016). As many repositories have not yet adopted generally accepted standards, it can be difficult for researchers to choose the right repository for their dataset (Dobratz et al. 2010).

Several organisations, including funding agencies, academic publishers, and data organisations provide researchers with lists of supported or recommended repositories, e.g. BioSharing (McQuilton et al. 2016). These lists vary in length, in the number and type of repositories they list, and in their selection criteria for recommendation. In addition, recommendations for data and data sharing are emerging, such as

the FAIR Data Principles: guidelines to establish a common ground for all data to be *Findable*, *Accessible*, *Interoperable*, and *Reusable* (Wilkinson et al. 2016). Some data repositories are beginning to incorporate the FAIR principles into their policies, such as the UK Data Service (2016) and several funders such as the EU Horizon 2020 program and the NIH (European Commission 2016; NIH Data Science 2016). Lists of recommended repositories and guidelines such as these can help researchers decide how and where to store and share their data.

Next to lists of recommended repositories, there are a number of schemes which specifically certify the quality of data repositories. One of the first of these certification schemes is the Data Seal of Approval (DSA), with an objective 'to safeguard data, to ensure high quality and to guide reliable management of data for the future without requiring the implementation of new standards, regulations or high costs' (DSA n.d. a). Building upon the DSA certification, but with more elaborate and detailed guidelines (Dillo & De Leeuw 2015), is the Network of Expertise in Long-Term Storage of Digital Resources (NESTOR) and the ISO 16363 standard/Trusted Data Repository (TDR). DSA, NESTOR, and TDR form a three-step framework for data repository certification (Dillo & De Leeuw 2015). The ICSU-WDS membership incorporates guidelines from DSA, NESTOR and Trustworthy Repositories Audit & Certification (TRAC), among others, for its data repository framework (ICSU-WDS 2012). Furthermore, the TRAC guidelines were used as a basis for the ISO 16363/TDR guidelines (CCSDS 2011).

Given the multitude of recommendations and certification schemes, we set out to map the current landscape to compare criteria and analyse which repositories are recommended and certified by different parties. This paper is structured as follows: first, we investigate which repositories have been recommended and certified by different organizations. Next, we provide an analysis of the criteria used by organisations recommending repositories and the criteria used by certification schemes, and derive a set of shared criteria for recommendation and certification. Lastly, we explore what this tells us about the overlap between recommendations and certifications.

Methods
1. Lists of repositories
1.1 Recommended repositories
To examine which repositories are being recommended, we looked at the recommendations of 17 different organisations, including academic publishers, funding agencies, and data organisations. These lists of recommended repositories are all the available recommendation lists currently found on the BioSharing website under the Recommendations tab (BioSharing n.d. a), and in a web search by using the term "recommended data repositories". These are lists compiled by the American Geophysical Union (AGU n.d.), BBSRC (BBSRC n.d.), BioSharing (BioSharing n.d. b), COPDESS (COPDESS n.d.), DataMed (DataMed n.d.), Elsevier (Elsevier n.d.), EMBO Press (EMBO Press n.d.), F1000Research (F1000Research n.d.), GigaScience (GigaScience n.d.), NIH (NIH n.d.), PLOS (PLOS n.d.), Scientific Data (Scientific Data n.d.), Springer Nature/BioMed Central (both share the same list) (Springer n.d.), Web of Science (Web of Science n.d.), Wellcome Trust (Wellcome Trust n.d. b), and Wiley (Wiley n.d.). All lists, including links to the online lists, were compiled into one list to compare recommendations (http://dx.doi.org/10.17632/zx2kcyvvwm.1). Not all data repositories indexed by the Web of Science's Data Citation Index (DCI) were included as there is no publicly available list with all repositories indexed by the DCI, so retrieval of recommended repositories was done through an individual search. The repositories indexed by Re3Data were not included in our list of recommended repositories as Re3data functions as "a global registry of research data repositories" (Re3Data n.d.) and thus does not recommend repositories. However, Re3Data was used to verify the repository's status, persistent identifiers, and obtained certifications.

1.2 Certified repositories
For our analysis of data repository certification schemes, we examined five certification schemes. These were the DSA, ICSU-WDS, NESTOR, TRAC, and ISO 16363/TDR. These schemes were chosen due to being used for certification (DSA, ICSU-WDS, NESTOR and ISO 16363/TDR) or as a self-assessment check for repositories (TRAC). Both the DSA (DSA n.d. b), ICSU-WDS (ICSU-WDS n.d.) and NESTOR (NESTOR Seal n.d.) provide lists of certified repositories on their respective websites. ISO 16363/TDR certification has not yet been awarded (Larrimer 2016). We consulted the websites of the DSA, ICSU-WDS, NESTOR, TRAC and ISO 16363/TDR certification schemes to see which repositories they certified. The results were compiled into one list (https://doi.org/10.17632/zx2kcyvvwm.1).
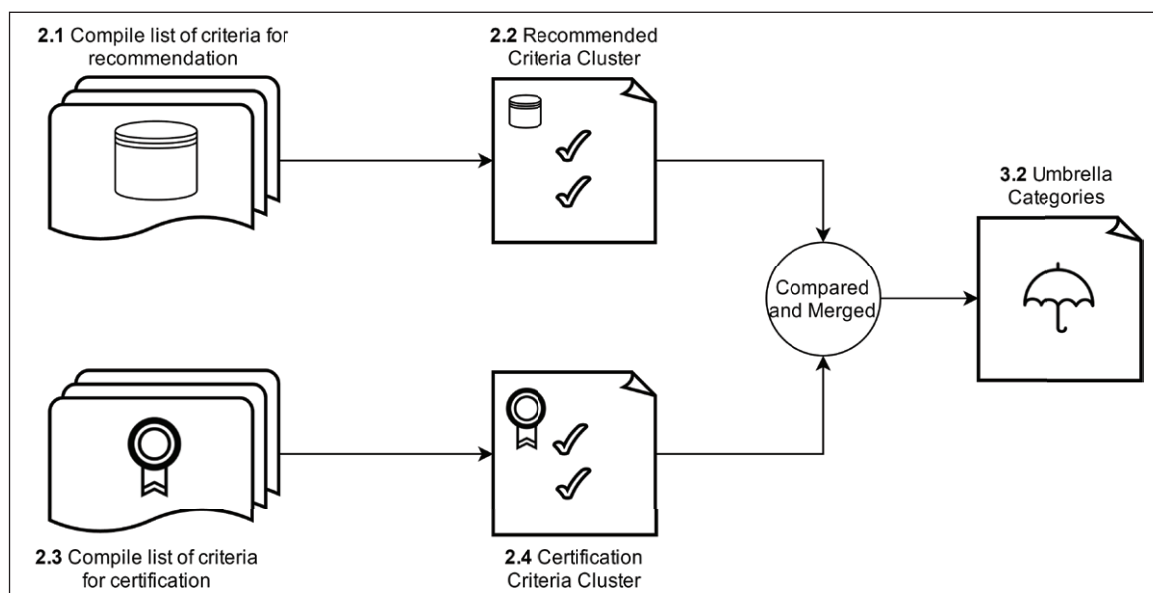
Figure 1: Flowchart of the methodology used to create the umbrella categories.

After composing the list of recommended repositories, we investigated which criteria are being used to determine a recommendation or certification, and whether an overlap exists between recommended and certified repositories, and the criteria used (**Figure 1**).

> 2.1 Compiled list of criteria for recommendation
> 2.2 Clustered criteria into the "Recommended Criteria Cluster"
> 2.3 Compiled list of criteria for certification
> 2.4 Clustered criteria into the "Certification Criteria Cluster"
> 3.2 Compared and merged steps 2.2 and 2.4 to create the umbrella categories

These steps will be discussed in turn.

2. Criteria used for recommendation and certification

2.1 Criteria for recommendation

To understand the motivation behind specific recommendations, we looked at the organisations' selection criteria for their lists of recommended repositories. Four of the 17 organisations supplied criteria for this online alongside their lists: BioCADDIE, F1000Research, Scientific Data/Springer Nature (SD/SN), and Web of Science (WoS). The Research Data Alliance's (RDA) criteria for recommended repositories were also included in this analysis. Although the RDA does not maintain a list of recommended repositories, we included their criteria to balance out the weight between the number of organisations that recommend repositories, and the number of organisations that provide criteria for certification. The criteria of the five organisations were then compiled into one list (https://doi.org/10.17632/zx2kcyvvwm.1).

2.2 Recommended Criteria Cluster

We categorized the criteria into 15 subheadings: Recognition, Mission, Transparency, Certification, Interface, Legal, Access, Structure, Retrievability, Preservation/Persistence, Curation, Persistent Identifier, Citability, Language, and Diversity of Data. These subheadings were derived from recurring and shared subjects throughout the different criteria lists. We then filtered out repetitions, or criteria unique to one organisation, namely Language and Diversity of Data, to create the "Recommended Criteria Cluster".

2.3 Criteria for certification

We consulted relevant websites to obtain the criteria used by the DSA, ICSU-WDS, NESTOR, TRAC and ISO 16363/TDR certification schemes and compiled a list of all certification criteria of the five schemes. In the case of the DSA (DSA n.d. c), ICSU-WDS (ICSU-WDS 2012) and NESTOR (NESTOR 2012), these were found

through their respective websites. The criteria for TRAC were found through the website of the Center for Research Libraries (CRL)(CRL and OCLC 2007), and the criteria for the ISO 16363/TDR were found through the Primary Trustworthy Digital Repository Authorisation Body, on the website of the Consultative Committee for Space Systems (CCSDS 2011).

2.4 Certification Criteria Cluster

The criteria found were categorized into 14 subheadings: Recognition, Mission, Transparency, Certification, Interface, Legal, Access, Indexation, Structure, Retrievability, Preservation/Persistence, Curation, Persistent Identifier, and Citability. These subheadings were derived from recurring and shared subjects throughout the different lists. Criteria that did not match any of the 14 subheadings, due to being too specific for one scheme, were categorized under a "miscellaneous" subheading. These criteria were then reorganized and filtered to remove repetitions and criteria unique to one certification scheme into 11 reworded categories: Community, Mission, Providence, Organisation, Technical Structure, Legal and Contractual Compliance, Accessibility, Data Quality, Retrievability, Responsiveness, and Preservation. We named this list the "Certification Criteria Cluster"

3. Overlap between recommended and certified repositories
3.1 Recommended versus certified repositories

To see whether there was overlap between the lists of recommended and certified repositories, we gathered the lists of repositories issued by the certifying organisations, and compiled these results into one list (available here: https://doi.org/10.17632/zx2kcyvvwm.1). We then calculated the number of times a repository was recommended by the different organizations as well as the percentage of recommended repositories with and without certification.

3.2 Criteria of recommendation and certification

We compared the Recommended Criteria Cluster and the Certification Criteria Cluster by looking at commonalities and recurrences between the two sets of broader headings and its constituents. We matched headings such as Community together with Recognition, Access with Accessibility, and Technical Structure with Interface, to derive a higher-order cluster of categories, which we are calling "Umbrella Categories". This was done to create broader terms that are relevant to both recommendation and certification criteria, to further ease the process of qualifying repositories. In the process of matching terms, those that were identical were left as is, e.g. Mission, and Retrievability.

3.3 Repositories by discipline

Next, we classified every recommended repository into one of five disciplines, to analyse if a relationship exists between certification, recommendation, and discipline. The disciplines we identified – General/Interdisciplinary, Health and Medicine, Life Sciences, Physical Sciences, and Social Sciences and Economics – were based on the focus disciplines provided on the websites of Scientific Data/Springer Nature, PLOS, Elsevier, and Web of Science, as these organisations provided comparable lists of disciplines.

Results
1. Lists of repositories

The lists of recommended repositories of all 17 organizations were compiled into one list together with the lists of certified repositories (https://doi.org/10.17632/zx2kcyvvwm.1) to allow for further analysis.

2. Criteria used for recommendation and certification

Based on the list of criteria that were used by the different organizations to recommend repositories, criteria were reorganized into 15 subheadings. After initial creation of the subheadings, we found some repetitions or scheme-specific criteria which we refined into 13 subheadings. The resulting list is the "Recommended Criteria Cluster" (Supplementary Table 1).

Similarly, based on the list of criteria of certification schemes, we were able to create 14 subheadings. These were re-examined and after refining, 11 subheadings remained. These resulting 11 subheadings and their criteria form the "Certification Criteria Cluster" (Supplementary Table 2).

3. Overlap between recommendation and certification

In total, we found 242 repositories that were recommended by publishers, funding agencies, and/or community organisations, and the distribution of recommendations is depicted in **Figure 2**. A majority (88) were only recommended by a single organisation; ArrayExpress was mentioned most frequently, being recommended by 12 out of 17 organisations.

When we look at recommended repositories with certification, only 13 out of 242 recommended repositories had any kind of certification: six repositories were certified by the DSA; six by the ICSU-WDS; none by NESTOR, TRAC, or TDR (**Figure 3**). Only one repository was certified by both the DSA and ICSU-WDS: the Inter-University Consortium for Political and Social Research (ICPSR). Of the 50 repositories that were recommended most often, only PANGAEA was certified.

To arrive at broadly applicable terms we grouped the criteria and the broader categories of the Recommended Criteria Cluster and the Certification Criteria Cluster into seven recurring umbrella categories, listed in **Table 1**.

When we looked at the discipline that each recommended repository covers, we found that that the most common focus discipline is Life Sciences with 115 repositories, followed by Physical Sciences with 80, Health and Medicine with 26, General/Interdisciplinary with 12, and Social Sciences and Economics with 9 (https://doi.org/10.17632/zx2kcyvvwm.1).

Discussion

In the analyses of criteria, we identified seven categories for criteria based on the shared aspects between the criteria for recommendation and the criteria for certification. These categories give an indication of the common requirements observed in this study. These suggest that a repository needs:

1) To have an explicit mission statement;
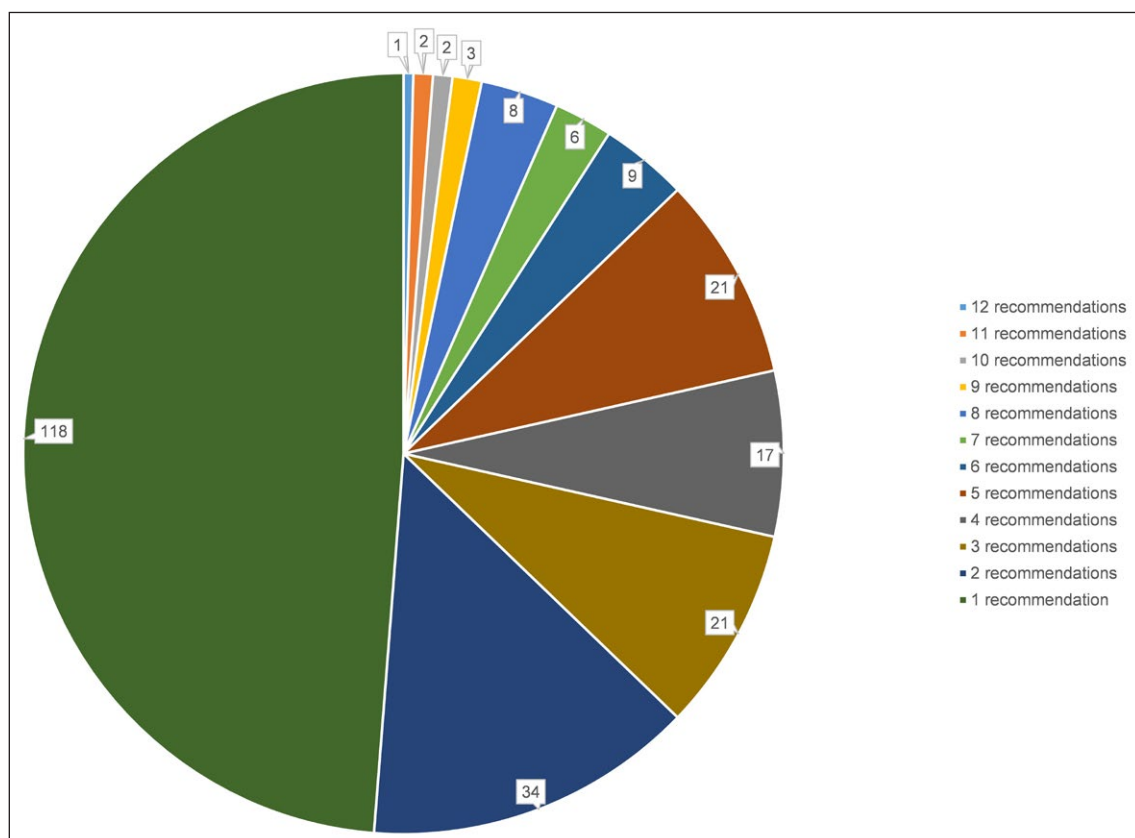2) To have a defined user community;

Figure 2: Pie-chart showing the number of repositories that are recommended by 1 to 17 organisations. The highest number of recommendations received is 12.
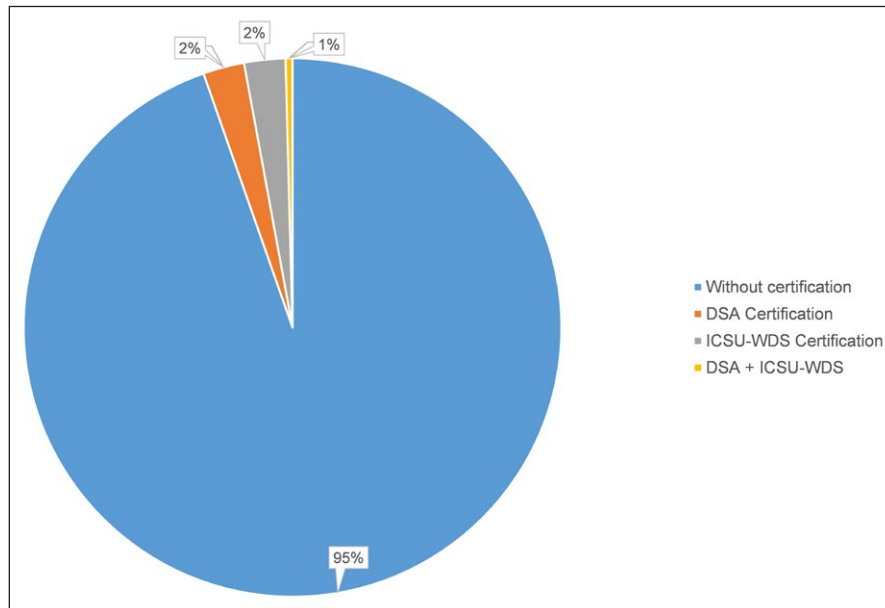
Figure 3: Pie-chart showing the percentage of recommended repositories that have obtained one of the following certifications: DSA, ICSU-WDS, and both DSA and ICSU-WDS.

Table 1: Table listing 7 common terms, referred to as "umbrella categories", based on a comparison between the Recommended Criteria Cluster and the Certification Criteria Cluster. The second column describes the shared meaning of the umbrella category, followed by differences in characteristics of recommended repository criteria and repository certification scheme criteria.

Umbrella Categories	Shared	Recommended Repository Criteria	Repository Certification Scheme Criteria
Mission	Explicit mission statement in providing long-term responsibility, persistence, and management of data(sets)		
Community/ Recognition		Evidence of use by downloads or citations from an identifiable and active user community	Understand and meet the needs of the designated and defined target community
Legal and Contractual Compliance	Repository operates within a legal framework/Ensures compliance with legal regulations	When applicable, have contractual regulations governing the protection of human subjects	Contracts and agreements maintained with relevant parties on relevant subjects
Access/ Accessibility	Public access to the scientific/ repository designated community	Anonymous referees (including peer-reviewers) have access to the data before public release as indicated by policies	
Technical Structure/ Interface		The software system supports data organisation and searchability by both humans and computers. The interface is intuitive and mobile user-friendly	The technical (infra)structure is appropriate, protective, and secure
Retrievability	Data need to have enough metadata. All data receive a persistent identifier		
Preservation	Long-term and formal preservation/succession plan for the data, even if the repository ceases to exist	If the data are retracted, the persistent identifier needs to be maintained	Preservation of data information properties and metadata

3) To operate within a legal framework;
4) To be accessible to the designated user community;
5) To have an adequate technical structure;
6) To ensure data retrievability;
7) To ensure long-term preservation of data.

In comparing lists of recommended and certified repositories, we found a strong discrepancy: only 13 out of the 242 repositories that were recommended have obtained any form of certification. This gap between recommendation and certification can also be found in the fact that of the 50 most often recommended repositories, only one is certified. There are several possible explanations for the existence of a gap between recommended and certified repositories.

Firstly, repository certifications find their origins in specific disciplines/domains. For instance, the DSA has a background in social sciences and humanities, NESTOR is formed by a network of museums, archives and libraries, and the ICSU-WDS mainly operates in the earth and space science domain (Dillo & De Leeuw 2015: 235–238). When looking at the different repository domains, we found that fifty of the most recommended repositories are linked to the natural sciences (life/physical/health sciences); exceptions in the top 50 are three interdisciplinary repositories. This depicts a contrast between repository domains and certifications' target domains. Repositories operating within the life sciences domain might be less aware of certifications while conversely, certifiers might be tailoring standards that are less focused on the requirements for life science repositories.

A second reason might be the dynamics in the field of data sharing; guidelines and practices are constantly evolving and being evaluated. While data repositories such as FlyBase have been active since the early 90s (Gelbart et al. 1997), certification schemes tend to be much more recent. The DSA is one of the oldest certifications, created in 2005 and becoming an internationally applicable scheme in 2009 (Dillo & De Leeuw 2015: 230). Changing data sharing dynamics are also reflected in the collaboration between certifying organisations and the merger of different guidelines. For instance, the partnership between the DSA and ICSU-WDS has brought together two certification schemes from diverse backgrounds and disciplines (Dillo & De Leeuw 2015: 237). Today, these schemes are still being developed and improved.

The only certifications that have been obtained by the repositories in our analysis are DSA and ICSU-WDS. NESTOR has certified two repositories at the time of writing, none of which were included in the examined recommendation lists. Both TRAC and ISO 16363/TDR have not officially certified repositories. In the case of the TRAC guidelines, this is because they are 'meant primarily for those responsible for auditing digital repositories [. . .] seeking objective measurement of the trustworthiness of their repository' (CCSDS 2011: 1), and are intended as a self-assessment check for repositories. The ISO 16363/TDR is a follow-up based on TRAC: it is an audit standard supported by the Primary Trustworthy Digital Repository Authorisation Body. Next to being considered a detailed way to evaluate a digital repository together with DSA and NESTOR (Dillo & De Leeuw 2015: 235), repositories can be ISO 16363/TDR accredited by the ANSI-ASQ National Accreditation Board, although so far this has not yet happened (Larrimer 2016).

Another possible cause of the gap between certification and recommendation is that most organisations do not ask repositories for certification. From the five organisations used to create our Recommended Criteria Cluster analysis, only the RDA requires repositories to have a certification. The RDA states that a "trustworthy" repository is one 'that untertake [sic] regularly quality assessments successfully such as Data Seal of Approval/World Data Systems' (RDA). Organisations not asking for certification might provide repositories with little incentive to work towards the requirements of certification schemes.

A limitation of this study is that we clustered the lists of all 17 organisations under the header of recommended repositories, but the way in which this recommendation happens differs. In some cases repositories are referred to by these lists as "supported" or "approved" (Elsevier, F1000Research), encouraged for their ability to make data reusable (NIH), meeting access, preservation, and stability requirements (Scientific Data/Springer Nature), having received funding or investment (BBSRC, Wellcome), or the list contains generally recognised repositories (PLOS). This can explain why overlap between the lists is limited, in certain cases, because different considerations might have played a role, depending on the intention of the organisation. We note that there might be additional lists available that were not found using our search methodology.

Conclusion

The umbrella categories which we identified for data repositories indicate that publishers, community organisations and certification schemes largely agree on quality. Yet this is not reflected in the relationship between recommended and certified repositories. Out of the entire list, less than 6% of recommended repositories obtained some form of certification. Programmes such as the Horizon 2020 programme of the European Union tell their grantees that when choosing a repository '[p]reference should be given to certified repositories which support open access where possible' (European Commission 2016), increasing the focus on certified repositories. This focus could create an incentive for repositories, in particular uncertified recommended repositories, to become certified.

To ensure certification schemes are known to repositories, we suggest domains collaborate in creating or maintaining certification schemes. While we see the value in having different levels of certification, the recent DSA – WDS partnership provides a good example of a collaboration between domains. If repositories and certification schemes work together towards common standards, this can provide clarity and improve data management and quality. It would improve the data landscape if collectively, we make sure that recommended repositories are certified, and certifications are obtained by repositories that are being recommended.

Further research into this area could involve looking at whether the criteria for repository certification have actually contributed to better data. It would be of interest to study if certification leads to more "FAIR" data, and whether data in a certified repository is more highly used and cited. Another possible topic of study is whether these criteria can be applied at the level of the dataset. This could also lead to the development of (semi-)automated tools to check whether datasets comply with the certification standards, before or during the process of submission of datasets to repositories. Our umbrella categories might help to develop the categories that datasets are judged against.

As a final conclusion, we suggest that the seven common umbrella categories we identified could form a common ground for the standardisation of data repository requirements. These categories are multifaceted in meaning and could therefore support and improve data management, data quality, and transparency of the services provided by the data repository. Our results complement previous work by Dabratz et al., 2010 in which she concludes that general standards for repositories are complex, standards are used as guidelines instead, and that there is a need for a specific standard. Our umbrella categories could provide the basis for this standard. Therefore, we suggest that researchers take these criteria into account when choosing a repository, and suggest that recommending organizations consider converging on these standards.

Additional Files

The additional files for this article can be found as follows:

- **Supplementary Table 1.** Table "Recommended Criteria Cluster" showing 13 main criteria used by organizations recommending repositories, with separate individual characteristics for each. The occurrence of criteria in the information provided by organizations is represented by an "X". DOI: https:/doi.org/10.5334/dsj-2017-042.s1
- **Supplementary Table 2.** Table "Certified Criteria Cluster" showing 11 main criteria used by organizations certifying repositories, with separate individual characteristics for each. The occurrence of criteria in the information provided by organizations is represented by an "X". DOI: https:/doi.org/10.5334/dsj-2017-042.s2

Competing Interests

The authors are affiliated with one of the organisations keeping a list of supported repositories.

Authors Contribution

Sean Edward Husen and Zoë G. de Wilde have contributed equally to this paper.

References

Assante, M, et al. 2016 Are Scientific Data Repositories Coping with Research Data Publishing? *Data Science Journal*, 15(6): 1–24. DOI: https://doi.org/10.5334/dsj-2016-006

BBSRC (n.d.) Resources. Available at: http://www.bbsrc.ac.uk/funding/apply/application-guidance/justification-resources/resources/ (Accessed January 27, 2017).

BioSharing (n.d.) a Recommendations. Available at: https://biosharing.org/recommendations/ (Accessed January 18, 2017).

BioSharing (n.d.) b Databases. Available at: https://biosharing.org/databases/?q=&selected_facets=recommended:true (Accessed January 27, 2017).

Borgman, C L 2012 The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6): 1059–1078. DOI: https://doi.wiley.com/10.1002/asi.22634

CCSDS 2011 Audit and Certification of Trustworthy Digital Repositories. Available at: https://public.ccsds.org/pubs/652x0m1.pdf (Accessed: 28 October 2016).

COPDESS (n.d.) Search. Available at: https://copdessdirectory.osf.io/search/ (Accessed January 27, 2017).

CRL and **OCLC** 2007 Trustworthy Repositories Audit & Certification: Criteria and Checklist. CRL and OCLC. Available at: http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf (Accessed April 7, 2017).

DataMed (n.d.) Repository List. Available at: https://datamed.org/repository_list.php (Accessed January 27, 2017).

Dillo, I and **De Leeuw, L** 2015 Ten Years Back, Five Years Forward: The Data Seal of Approval. *International Journal of Digital Curation*, 10(1). DOI: https://doi.org/10.2218/ijdc.v10i1.363

Dobratz, S, et al. 2010 The Use of Quality Management Standards in Trustworthy Digital Archives. *International Journal of Digital Curation*, 5(1):46–63. DOI: https://doi.org/10.2218/ijdc.v5i1.143

DSA (n.d.) a About. Available at: http://www.datasealofapproval.org/en/information/about/ (Accessed January 18, 2017).

DSA (n.d.) b Assessment. Available at: http://www.datasealofapproval.org/en/assessment/ (Accessed February 20, 2017).

DSA (n.d.) c The Core Trustworthy Data Repository Requirements. Available at: https://www.datasealofapproval.org/en/information/requirements/ (Accessed April 7, 2017).

Elsevier (n.d.) Supported Data Repositories. Available at: https://www.elsevier.com/books-and-journals/enrichments/data-base-linking/supported-data-repositories (Accessed January 27, 2017).

EMBO Press (n.d.) Author Guidelines. Available at: http://msb.embopress.org/authorguide#datadeposition (Accessed January 27, 2017).

European Commission 2016 H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. Available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (Accessed October 27, 2016).

F1000Research (n.d.) Data guidelines. Available at: https://f1000research.com/for-authors/data-guidelines (Accessed January 27, 2017).

Gelbart, W M, et al. 1997 FlyBase: A Drosophila database, *Nucleic Acids Research*, 25(1): 63–66. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC146418/pdf/250063.pdf (Accessed October 21, 2016). DOI: https://doi.org/10.1093/nar/25.1.63

GigaScience (n.d.) Editorial Policies and Reporting Standards. Available at: https://academic.oup.com/gigascience/pages/editorial_policies_and_reporting_standards (Accessed January 27, 2017).

ICSU-WDS 2012 Certification of WDS Members. Available at: https://www.icsu-wds.org/files/wds-certification-summary-11-june-2012.pdf (Accessed: 28 October 2016).

ICSU-WDS (n.d.) Membership. Available at: https://www.icsu-wds.org/community/membership/regular-members (Accessed February 20, 2017).

Larrimer, N 2016 ANAB Heads Up. Available at: http://anab.org/media/59172/hu327.pdf (Accessed October 28, 2016).

Mayernik, M S, et al. 2014 Peer Review of Datasets: When, Why, and How. *Bulletin of the American Meteorological Society*, 96(2): 191–201. DOI: https://doi.org/10.1175/BAMS-D-13-00083.1

McQuilton, P, et al. 2016 BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences. *Database (Oxford)* 2016(2016): baw075. DOI: https://doi.org/10.1093/database/baw075

Merson, L, Gaye, O and **Guerin, P J** 2016 Avoiding Data Dumpsters — Toward Equitable and Useful Data Sharing. *New England Journal of Medicine*, 374(25): 2414–2415. DOI: https://doi.org/10.1056/NEJMp1605148

NESTOR 2012 Certification Working Group. Explanatory notes on the nestor Seal for Trustworthy Digital Archives. Available at: http://d-nb.info/1047613859/34 (Accessed on 7 April, 2017).

NESTOR Seal (n.d.) Available at: http://www.langzeitarchivierung.de/Subsites/nestor/EN/Siegel/siegel_node.html (Accessed February 20, 2017).

NIH (n.d.) NIH Data Sharing Repositories. Available at: https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html (Accessed January 27, 2017).

NIH 2015 NIH Grants Policy Statement. NIH. Available at: http://grants.nih.gov/grants/policy/nihgps/nihgps.pdf.

NIH Data Science 2016 NIH Commons Overview, Framework & Pilots – Version 1. NIH Data Science at NIH. Available at: https://datascience.nih.gov/commons.

Piwowar, H A, Day, R S and **Fridsma, D B** 2007 Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE*, 2(3): e308. DOI: https://doi.org/10.1371/journal.pone.0000308

Piwowar, H A and **Vision, T J** 2013 Data reuse and the open data citation advantage. *PeerJ*, 1: e175. DOI: https://doi.org/10.7717/peerj.175

PLOS (n.d.) Data Availability. Available at: http://journals.plos.org/plosbiology/s/data-availability (Accessed January 27, 2017).

RDA, Repository Bundle (n.d.) Research Data Alliance. Available at: https://www.rd-alliance.org/group/data-fabric-ig/wiki/repository-bundle.html (Accessed January 27, 2017).

Re3Data (n.d.) Re3Data.org|About. Available at: http://www.re3data.org/about (Accessed January 27, 2017).

Scientific Data (n.d.) Recommended Data Repositories : Scientific Data. Available at: http://www.nature.com/sdata/policies/repositories (Accessed January 27, 2017).

Springer (n.d.) Data policy – repositories. Available at: http://www.springernature.com/gp/group/data-policy/repositories (Accessed January 27, 2017).

UK Data Service 2016 The "FAIR" principles for scientific data management. UK Data Service. Available at: https://www.ukdataservice.ac.uk/news-and-events/newsitem/?id=4615 (Accessed October 28, 2016).

Web of Science (n.d.) Master Data Repository List. Available at: http://wokinfo.com/cgi-bin/dci/search.cgi (Accessed January 27, 2017).

Wellcome Trust (n.d.) a Policy on data management and sharing. *Wellcome.* Available at: https://wellcome.ac.uk/funding/managing-grant/policy-data-management-and-sharing (Accessed January 27, 2017).

Wellcome Trust (n.d.) b Data repositories and database resources. Available at: https://wellcome.ac.uk/funding/managing-grant/data-repositories-and-database-resources (Accessed January 27, 2017).

Wilkinson, M D, et al. 2016 The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3:160018. DOI: https://doi.org/10.1038/sdata.2016.18