cessda

CESSDA Work Plan 2020

WPT: Widening and Journals Outreach 2020

Assessment of Service Provider Capacities

Deliverable 2: Report

Document info

Dissemination Level	PU
Due Date of Deliverable	31/10/20
Actual Submission Date	30/10/20
Type	Report
Approval Status	Approved by CESSDA Training Working Group Leader Irena Vipavc Brvar
Version	V3.0
Number of Pages	p.1 – p.23
DOI	10.5281/zenodo.5554432

Version history

Version	Date	Comment	Revised by
0.0	Sep-Oct 2020	Co-edited electronic document	All authors in list below
0.1	October 2020	Draft in template to all co-authors	Serafeim Alvanides
0.2	October 2020	Edits and comments to v0.1	ADP team
0.3	October 2020	Edits and comments to v0.2	CROSSDA team
0.4	October 2020	Edits and comments to v0.3	FORS team
0.5	October 2020	Edits and comments to v0.4	TARKI team
0.6	29.10.2020	Peer review	Jindřich Krejčí, CSDA
1.0	30.10.2020	Submission to CESSDA	Serafeim Alvanides
1.0MEdit	11.11.2020	Feedback from CESSDA	CESSDA MO
1.1	20.11.2020	Submission to CESSDA	Serafeim Alvanides
1.1WGL	21.01.2021	Feedback from CESSDA	CESSDA MO
2.0	24.02.2021	Submission to CESSDA	Serafeim Alvanides
2.0WGL	26.03.2021	Feedback from CESSDA	CESSDA MO
3.0	27.04.2021	Submission to CESSDA	Serafeim Alvanides

Author List

Organisation	Name	Contact information
GESIS, Germany	Serafeim Alvanides	Serafeim.Alvanides@gesis.org
ADP, Slovenia	Janez Štebe	janez.stebe@fdv.uni-lj.si
FORS, Switzerland	Emilie Morgan De Paula	emilie.morgandepaula@unil.ch
FORS, Switzerland	Brian Kleiner	brian.kleiner@fors.unil.ch
TARKI, Hungary	Peter Hegedus	peter.hegedus@tarki.hu
CROSSDA (FFZG), Croatia	Marijana Glavica	mglavica@ffzg.hr
CROSSDA (FFZG), Croatia	Irena Kranjec	ikranjec@ffzg.hr

Peer-review

Organisation	Name	Contact information
CSDA, Czech Republic	Jindřich Krejčí	jindrich.krejci@soc.cas.cz

Contents

Executive Summary

The overall goal of the "CESSDA Widening Activities and Journal Outreach 2020" Task 4 (henceforth referred to as *Task 4*) is to define and promote a collective position statement from CESSDA regarding archival services for scholarly journals. The first deliverable reported on journal practices, requirements and needs regarding timely and user-friendly accessibility of data from scholarly publications (Alvanides et al. 2021). The objective of this deliverable is to provide an assessment of CESSDA Service Providers' capacities, policies, and services for responding to the needs of journals with respect to access and medium/long term preservation of data and replication material. To this effect, an online survey targeting the current consortium of 22 CESSDA Members was conducted, with questions guided by the findings of the first deliverable. The survey sought to profile SPs in relation to the requirements expected by journals, to assess potential limitations encountered, to gather information on actual and planned collaboration with journals, to investigate potential support that SPs would benefit from and to gauge their views on CESSDA's involvement in coordinating such initiatives.

The first major finding is that SPs are well placed to offer long-term curated data collections, quality control, personalised expert curation support and links to publications, training on data management planning and promotional material related to data usage. These are all services at the core of the SPs' mission and they can benefit journals and authors who are focusing on established social surveys deposited with the archives, but also have the potential to encompass diverse research datasets. On the other hand, very few SPs offer or engage with replication support services. In addition, most SPs perceive the needs of journals in relation to the service already provided, such as PID and versioning of datasets, data usage information, citations tracking and links to publications. Again, these are important services delivered at a high standard, demonstrating SPs' awareness of the evolving landscape, but journal editors almost take these services for granted, while some social science disciplines (e.g. economics, psychology, occasionally sociology) are moving rapidly ahead in their thinking about what constitutes open science. Therefore, there is good potential for SPs (and by extension for CESSDA) to promote these services to journal editors and publishers, as expectations are gradually moving further in terms of what constitutes open science. SPs are aware of this shift, by identifying the complexities associated with the diversity of social science data, such as handling new data types, diverse datasets resulting from inter- or multidisciplinary research and dealing with sensitive data that require controlled or restricted access. The survey revealed that currently more than half SPs offer low or no-cost self-archiving, a service in high demand by researchers and journal editors, while one third have no plans to make it available in the near future. Echoing the views of journal editors in relation to their journals, the SPs also highlighted limitations with financial and technical resources in order to fully implement the data sharing and replicability policies envisaged. This situation leaves a void, but also an opportunity for fruitful collaboration between the interested parties.

The second major finding is that the SPs are overwhelmingly positive towards offering data sharing and replication services in collaboration with publishers and journals. Half of the SPs have already received data and associated materials linked to journal articles, mostly without major technical or policy issues, which sets a positive precedent for the other SPs. The majority of SPs responded positively to the suggestion that CESSDA should take on a coordinating role, highlighting the CESSDA "brand" and expertise as driving forces for involvement, listing benefits of collaboration in relation to knowledge exchange, good practice and economies of scale, and acknowledging the current expertise amongst CESSDA members. There is a potential opportunity cost from CESSDA not coordinating this action and the prospect that the current void may be filled by the journal publishers themselves, at the risk of not involving the national archives. This was also a concern raised by the journal editors in relation to understanding the policies on data protection, storage and sharing at the national and European levels. Therefore, CESSDA's coordination role should aim beyond data sharing and replicability to encompass reproducibility, trust and transparency of published research as the aim of SP collaboration with journals.

This report has completed an assessment of the CESSDA SP's capacities, policies, and services for responding to current and potential needs of journals regarding the preservation of data and replication material. There is a good understanding of the evolving landscape and willingness of SPs to develop additional services that fulfil the data sharing needs of journals, but also an expectation for CESSDA to become more involved and facilitate collaboration and sharing of expertise and good practice. The next objective of *Task 4* is to explore models for how CESSDA could coordinate SPs and to propose models for how SPs may collaborate in offering additional services to journals.

Abbreviations and Acronyms

CC	Creative Commons licenses
CESSDA	Consortium of European Social Science Data Archives
FAIR	Findability, Accessibility, Interoperability and Reusability
GDPR	General Data Protection Regulation
IJSRM	International Journal of Social Research Methodology
RDM	Research Data Management
SP(s)	Service Provider(s)

1. Background

There is heightened demand from research funding bodies for open access of academic articles published in scholarly journals and from the scientific community for more transparency and reproducibility of findings, including replication. This has resulted in an increasing number of publishers offering additional services beyond open access of scholarly articles. By encouraging and, in some cases, requiring that authors share their data and code in order to publish their research, publishers and journals have become an important agent in the movement to improve the openness of data and the reproducibility of research. A number of highly ranked social science journals have implemented data policies and mandate sharing, led by journals in economics, followed by political science/international relations and psychology (Crosas et al. 2018) and other disciplines (Resnik et al. 2019, Rousi & Laakso 2020). This trend is increasing and presents a unique opportunity for the CESSDA SPs to diversify their services, thus catering for the evolving needs of publishers, journals and researchers in relation to open science.

The overall goal of *Task 4* is to define and promote a collective position statement from CESSDA regarding data sharing services offered to journals. The first report (Alvanides et al. 2021) provided a better understanding of journal practices, requirements and needs regarding timely and reader-friendly accessibility of data used in scholarly publications. The current report provides an assessment of CESSDA SP's capacities, policies, and services for responding to the needs of journals with respect to sharing and medium/long term preservation of data and related materials (such as related publications, instruments, documentation, code for analysis etc.). The document describes briefly the methodology used for primary data collection, the findings from the analysis of the SP survey and the steps forward in relation to the remaining two objectives and deliverables of *Task 4*.

2. Methodology and data collection

In order to achieve the objective of the second *Task 4* deliverable an online survey was designed and conducted, targeting the current consortium of 22 CESSDA Members. It is important to point out here that CESSDA SPs are at various stages of maturity, from long-established to relatively new archives, some of them still in early stages of development, and this is reflected in some of the responses. The survey questions were guided by the findings of the first report (Alvanides et al. 2021) with a focus on journal requirements and needs regarding the availability of data used in scientific publications. Following an assessment of the current journal practices and responses from editors, this survey was designed around two sections. The first part consisted of mostly closed (choices and ranking) questions to profile SPs in relation to requirements from publishers and journals as well as current and potential limitations that SPs may be encountering. The second part of the survey sought information on actual collaboration with journals (or planned collaboration in some instances), the type of support provided to journals and

authors, current challenges and potential support that SPs could benefit from. The survey concluded with questions on CESSDA's potential role in relation to such initiatives (open-ended question) and an offer to be contacted in relation to the next activity of *Task 4*.

The actual survey was developed online by CROSSDA in LimeSurvey and hosted by the University of Zagreb University Computing Centre (SRCE) in Croatia. The survey was launched on 21st August with an invitation posted to the *Service Providers' Forum* on the Basecamp platform. Initially the survey was planned to run until 3rd September 2020, but due to the summer holiday period and the low response rate, a decision was made to extend the survey period until 23rd September 2020 (Appendix B). This extension allowed for personalised contact with SPs, resulting in 19 responses from the 22 SPs (~86% response rate). The survey responses were downloaded and analysed using descriptive statistics and graphs in Excel, as well as thematic analysis of the open-ended questions.

3. Analysis: archive profiles in relation to journal requirements

Following an extension to the survey deadline, SPs from 19 countries responded (in alphabetic order): Austria, Belgium, Croatia, Czech Republic, Denmark, France, Germany, Greece, Hungary, Ireland, Netherlands, North Macedonia, Portugal, Serbia, Slovakia, Slovenia, Sweden, Switzerland and the United Kingdom. The opening question regarding the archive listed in selected repository registries and consideration for the future to be included in registries was answered as shown in Fig.1 below. The majority of responding SPs (12/19) are already listed in Re3ata with 3 also listed in FairSharing, while 7 of the SPs are not listed in any registry, but they would consider being included in one in the near future.

Figure 1: Listing of CESSDA SP archives in repository registries.
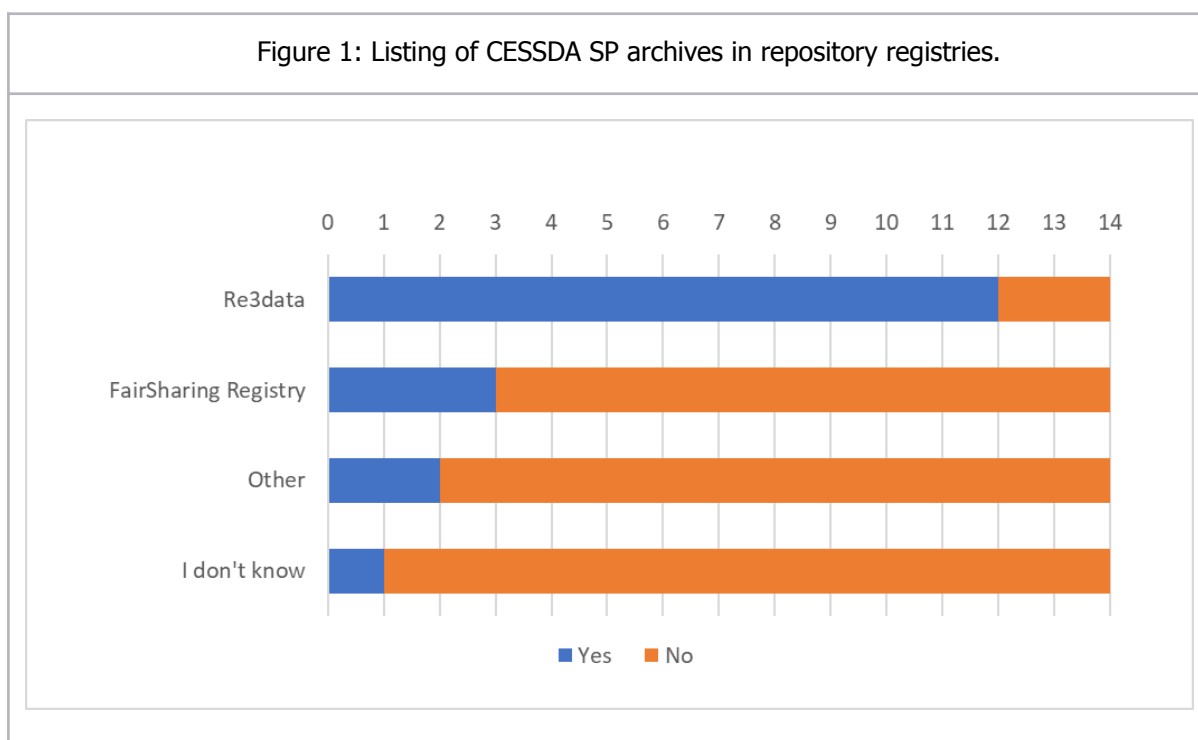
Moving on from repository registries, the next question was in relation to services that might be of relevance for data-related journal requirements currently available or planned at the SPs' data archive. As shown in Fig.2, almost all the SPs offer these services currently or they plan them for the near future:

- Long-term curated data collections (A),
- Quality control, e.g. data consistency, checks on data anonymisation, documentation completeness (C),
- Personalised expert curation support for data and metadata preparation and deposit (D),
- PID and versioning for datasets (E),
- Training, advice and active support for data management planning (O) and
- Promotion of data usage (e.g., via blog, newsletter, trainings for future data users etc.) (P).

The reason why some SPs are not currently offering long-term curating of data collections (A) and quality control (C) is that these are high cost services and not all SPs have enough resources or capacity to develop them yet.

A more mixed picture emerged in relation to services that are available for some and planned for other archives (Fig.2), such as:

- Relation in metadata to ORCID, or similar registries (F),
- Specific data usage information or citations tracking (G) and
- Links to publications (H) and Standard licences (L).

These indicate willingness to engage with such services, but potential limitations due to the more complex nature of creating relations and links for these services. Short-term self-archiving (B) presented an interesting pattern with more than half currently offering this service, and one third not making it available nor planning to in the near future. On the other hand, a significant number of services are neither currently available nor planned for the majority of archives (Fig.2), such as:

- Linking to related software (I),
- Access to data for peer review during embargo period (J),
- Support for 'double-blind' peer review of data during embargo (K),
- Pre-registration service (M) and
- Replication support service (N).

These are all more advanced and technically demanding services that would potentially satisfy the requirements of journals and researchers, but such services would also require both technical and scientific input in order to be implemented, possibly in collaboration with specific journals. It should also be noted that some of these services may simply be out of the scope of certain data archives. In addition, there was an open question offering the option to comment in more detail where one responding SP highlighted the lack of personnel

www.cessda.eu

or funding resources in order to implement future plans (e.g. "linking data to articles, bibliography and definitions, terms etc.") and another commented very positively on the Dataverse software for offering many of the above services.

Figure 2: SP services relevant to journal requirements available or planned in the near future.
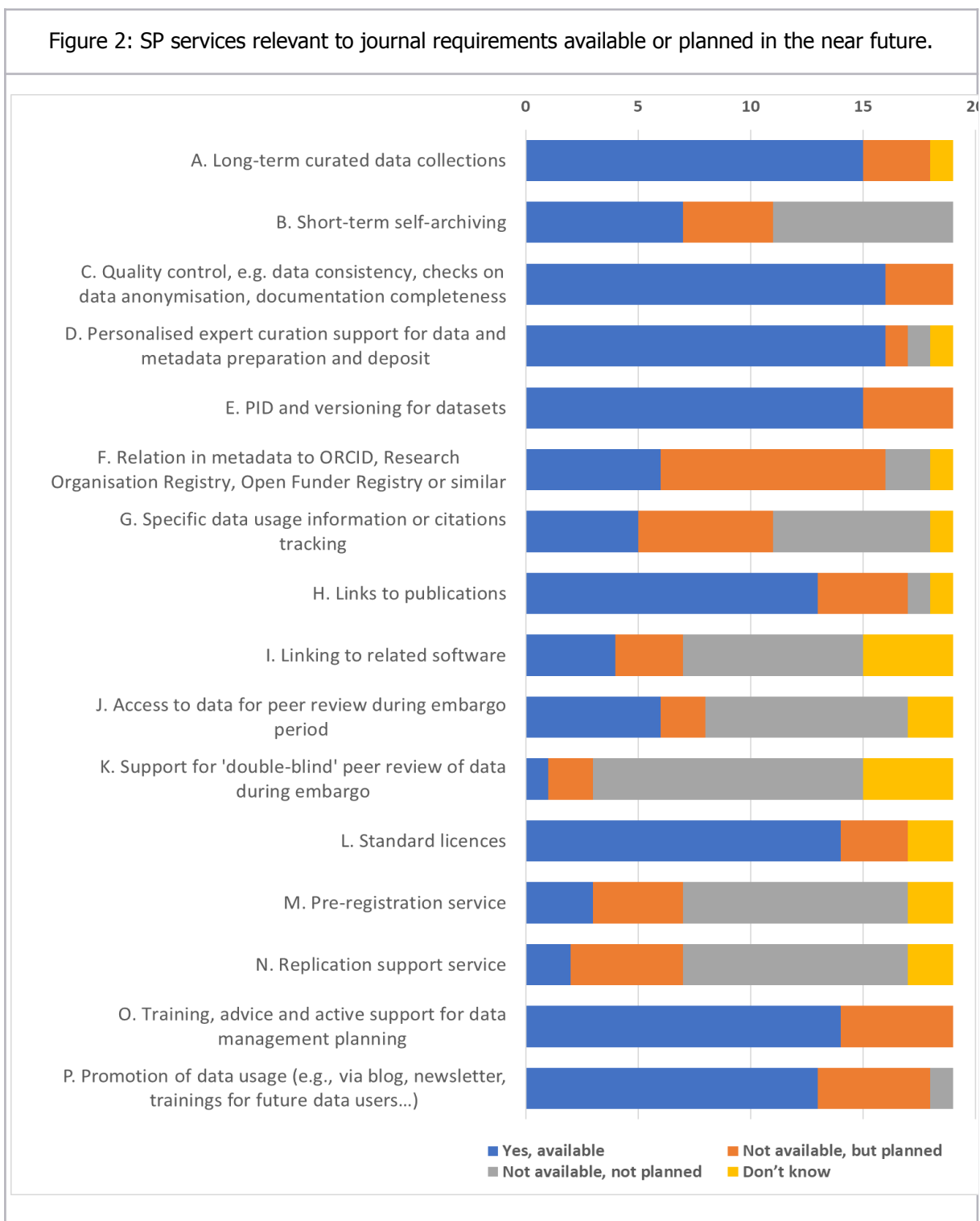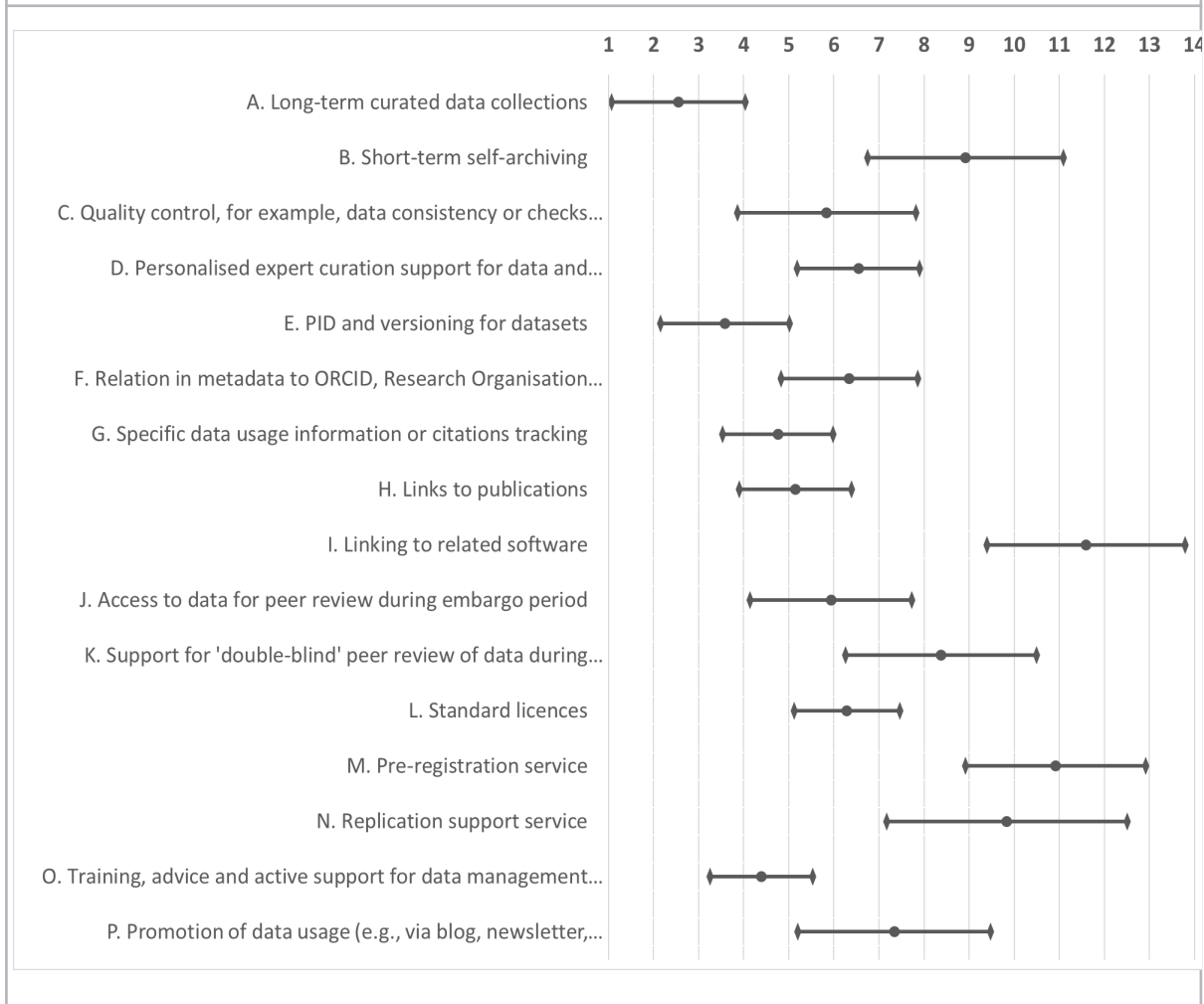
Asking respondents to rank the above 16 options in order of perceived importance to journal authors revealed some interesting patterns for the highest ranked services, as shown in Fig.3. The service considered most important with an average ranking of 2.5 as an incentive for authors to deposit data with archives was offering them Long-term curated data collections (A). This was followed by average rankings of 3.5 for PID and versioning of datasets (E) and 4.3 for Training, advice and active support for data management planning (O). The next two options with relatively high average rankings were Specific data usage information or citations tracking (G) with 4.7 and Links to publications (H) with 5.1. Comparing these 5 services to what is currently on offer by the archives in the previous question, reveals a reasonable match: four of the highest scored services in terms of importance: A (long-term curation), E (PID and versioning), O (data management planning support) and H (links to publications) are also already available or planned by the archives as discussed earlier (Fig.2). However, although option G (data usage/tracking) scored relatively high as a potential incentive for the authors, only a handful of SPs offer this service and a significant number are not planning to offer such services in the near future. This is also the case for Access to data for peer review during embargo period (J), which scored an average rank of 5.8, but very few SPs offer this service and most have no plans to offer it in the future. On the other hand, Quality control (C) is a service offered by most SPs and the remaining few are planning to offer it in the future (Fig.2), but with an average 5.8 do not necessarily rank it as one of the strongest incentives.

Continuing with the ranking of incentives to authors, three services were ranked on average around the middle of the scale: Promotion of data usage (P) with 7.3, Support for 'double-blind' peer review of data during embargo (K) with 8.3 and Short-term self-archiving (B) with 8.8 (Fig.3). Promotional activities (P) are offered or planned by almost all SPs, self-archiving (B) is currently offered or planned by more than half, but 'double-blind' peer review of data is not a current or future priority for SPs (Fig.2). Towards the end of the scale, Pre-registration (M) and Replication support (N) services generally ranked as low incentives for the authors to deposit data with average rankings 10.8 and 9.8 respectively (Fig.3). These rankings clearly reflect the fact that they are neither available nor planned for most of the SPs, as discussed earlier (Fig.2). The lowest average ranking was for Linking to related software (I), but with only six respondents ranking this option, the mean value 11.5, has wide confidence intervals.

Finally, one of the respondents provided comments in relation to Support for 'double-blind' peer review of data during embargo (K): "a dataset belonging to a publication that is documented, will always contain information about the people involved in the data collection; this does not necessarily mean that they are also the authors of the article, but there is a good chance they are, so this is very difficult if not impossible to guarantee". This is a valid point and reflects the middle-to-low average ranking of such a service.

Figure 3: SPs' ranking on importance of services as incentives for authors to deposit data.
Average rankings with 95% confidence intervals (1 is ranked as most important).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A. Long-term curated data collections														
B. Short-term self-archiving														
C. Quality control, for example, data consistency or checks…														
D. Personalised expert curation support for data and…														
E. PID and versioning for datasets														
F. Relation in metadata to ORCID, Research Organisation…														
G. Specific data usage information or citations tracking														
H. Links to publications														
I. Linking to related software														
J. Access to data for peer review during embargo period														
K. Support for 'double-blind' peer review of data during…														
L. Standard licences														
M. Pre-registration service														
N. Replication support service														
O. Training, advice and active support for data management…														
P. Promotion of data usage (e.g., via blog, newsletter,…														

The final question from the first part of the questionnaire was related to archives' limitations on accepting data and documentation with seven specific characteristics. As shown in Fig.4 for the majority of SPs, there were no limitations in relation to Data from projects not funded by specific funding sources or from the depositors' institutions (B), with Code/syntax for preparing data and/or reproducing analysis (G) in the second place. The highest (potential or strong) limitations were in relation to Types and formats of data outside the archives' scope of policy (E) alongside Data from disciplines other than social sciences (F). These two responses reflect the potentially unambiguous mission, scope and policy of respective archives focusing on social sciences. However, they can also be interpreted as potential mismatch between the services currently on offer and increasing trends towards inter- and multidisciplinary research generating "non-traditional" social science datasets. The next limitation with high responses (Fig.4) was datasets containing language other than English or official languages of the country the archive is located. With pan-European social surveys and new types of datasets on the increase (e.g. from social media), this can be a

very serious limitation in the context of SPs potentially offering archiving services to journals.

Figure 4: Limitations on accepting data and documentation.
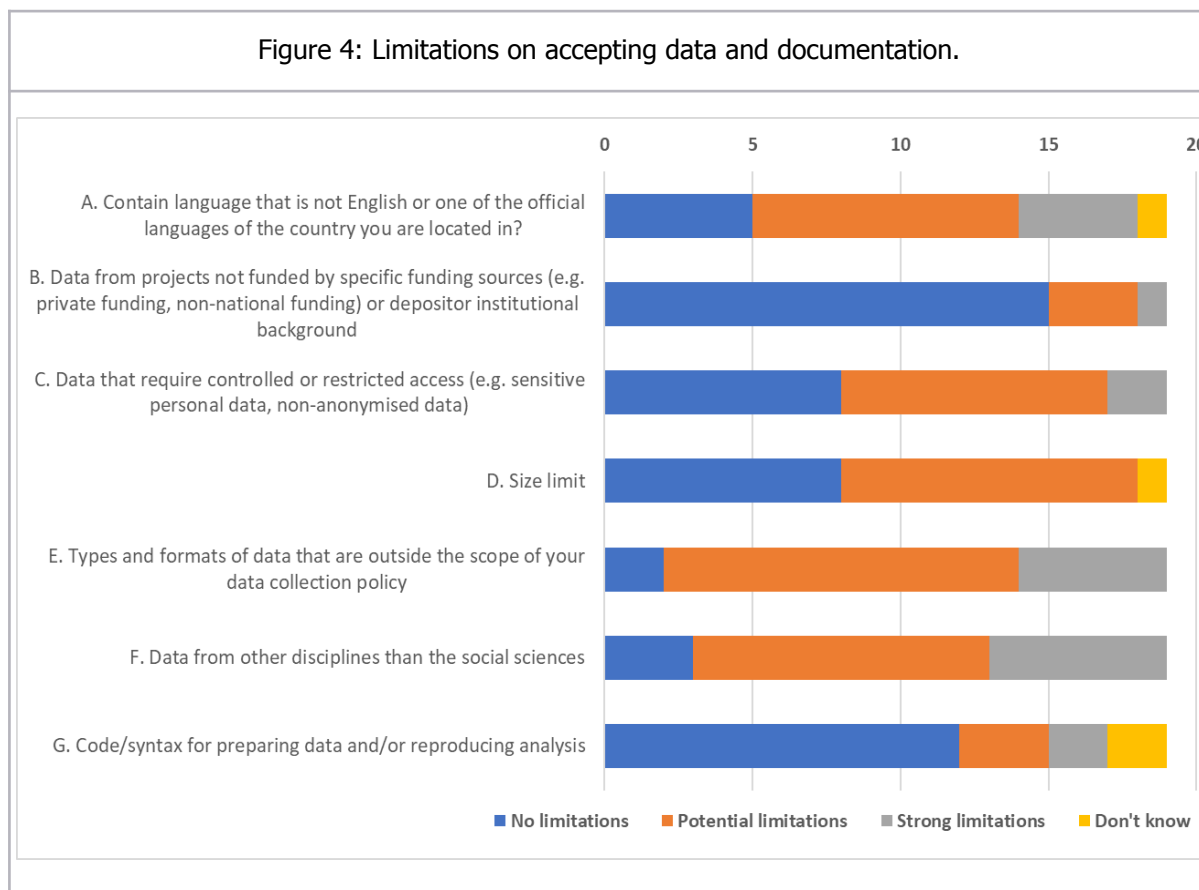
Some of the options divided the respondents between those stating no limitations and those perceiving potential/strong limitations. In particular, Size limit of datasets (D) and Data that require controlled or restricted access, such as sensitive personal data, non-anonymised data etc. (C), both revealing potential limitations with resources, storage space for the former and physical or online secure data centres for the latter. With increasing availability of very large datasets ("big data") and linked as well as georeferenced data being used extensively by social scientists, these limitations are likely to pose real challenges if the archives wish to offer archiving services to high impact journals.

4. Analysis: collaboration with journals

The second part of the questionnaire focused on current and planned collaborations of each archive with journals. The opening question was if the archives accept data and associated materials linked to journal articles with 15 out of 19 responses positive. The majority of those accepting (9/15) responded that they already have data and associated materials linked to journal articles deposited with them. Of those accepting data linked to journal

articles, two thirds (10/15) stated that they are not facing major technical or policy issues, while SPs facing challenges, responded as follows:

> "data based on secondary sources and data from foreign sources with no connection to [archive's country]"

> "we cannot accept all data linked to articles in a particular journal if the journal publishes articles in the fields other than social sciences"

> "datasets usually not published on the web, but journals asking this possibility so we have to add the dataset on to the archiving page"

> "it is not clear on what basis Journals include data archives in their list of preferred repositories"

The first two statements reflect concerns around mission and scope of the archive in terms of geographical coverage and disciplinary focus. The latter two statements reveal communication challenges with the journals in terms of what the journals' expectations are and how they implement their own data sharing policies.

Interestingly, the vast majority of those accepting data related to journals (13/15 would not need additional support in order to accept data linked to journal articles, while one responding SP explained that they would simply have to update their metadata schema. These responses and comments demonstrate that the 15 archives that currently accept data and associated materials linked to journal articles have also resolved most of the technical and policy issues, with the possible exception of potential improvements in communicating the scope and focus of the archives in relation to expectations from the journals.

Regarding formal collaboration with journals or publishers, almost half of the SPs (7/15) responded positively and six provided further details, evidencing various degrees of engagement:

- AUSSDA (Austria) provides archiving of journal data holdings, a platform for archiving replication data and agreements to inform and link authors to the repository, alongside consultancy and support in archiving options and RDM.
- PROGEDO (France) is a member of OPERA and has set up a recent collaboration with OpenEdition enabling them to accept data from OpenEdition journals.
- ADP (Slovenia) supports two journals (within a RDA node SI) with instructions for authors on citing data and recommendations to deposit data with trusted domain specific repositories. They provided advice on designing open data policies and agreed at level of support to the journals during the implementation process and to the authors once the policy is implemented.
- DANS (The Netherlands), seems to be the most active in collaborations having set up the *Research Data Journal for the Humanities and Social Sciences* in collaboration with Brill and acting as the backup-archive for Elsevier's Mendeley Data which

contains many data sets linked to publications. Their support is limited to linking to the publications of which the data are deposited in the archive.

- SND (Sweden) is on PLOS ONE's list of recommended repositories.
- UKDS (United Kingdom) collaborates with the journals *Scientific Data* and *Nature*, providing single blind peer review of data collections to support data articles.

Of the four responding SPs that do not currently accept data and associated materials linked to journal articles, two are not planning to go in this direction, in the foreseeable future. On the other hand, one SP is planning to start accepting data linked to journal articles when they complete their IT dissemination project and another SP is considering going this direction when the main operations of the recently established archive are fully functional.

Moving on to the question about CESSDA adopting a coordinated approach in supporting social science journals with the involvement of SPs, the vast majority of respondents (15/19) saw this as a positive way forward with 3 abstaining (Don't know) and only one archive expressing a more sceptical view of this suggestion. Thirteen (13) of the positively disposed respondents provided additional comments, elaborating on why CESSDA's involvement is seen as a good approach. The responses are presented in Table 1 (slightly edited to secure anonymity) and grouped into three themes, following thematic analysis. Some of the responding SPs provided more than one sentences, resulting in more than 13 bullet points in Table 1.

These comments clearly reflect the high esteem and positive attitude towards CESSDA's expertise, alongside an understanding of the promising potential afforded through collaboration in terms of knowledge exchange, good practice and economies of scale. The third theme indicates concerns raised as a result of CESSDA not becoming involved (e.g. potential lack of data transparency), and the opportunity costs with other publishers filling the void left by the lack of engagement. Interestingly, the final comment listed in Table 1 highlights the importance of accommodating continuous progress of SPs, depending on their stage of development. The single critical response received under the question "Why do you think that this would not be a good approach?", was in relation to CESSDA's geographical focus at the exclusion of English-speaking countries outside Europe. The suggestion from the critical response was that IASSIST[1] with its international remit may be better placed than CESSDA to take the lead on this.

Finally, in the space inviting respondents to share further comments, SODHA (Belgium) and MK DASS (Northern Macedonia) responded that they are very interested in archiving data for journals, but currently engaged with launching their archives. DANS (The Netherlands) also expressed strong interest in the activities of *Task 4* and highlighted the *Research Data Journal (RDJ)* as a potential model. The DANS respondent also pointed out that a distinction should be made between the deposit of original data (on which the research and article is

---

[1] IASSIST (International Association for Social Science Information Service and Technology) webpage: https://iassistdata.org/about [23 February 2021]

based) and "additional materials" (such as figures, tables and small spreadsheets), noting that the focus of CESSDA and national archives should be on archiving and preserving the original data sets related to the articles.

Table 1: SPs' views on a CESSDA coordinated approach towards supporting social science journals. Grouped into themes based on context (some respondents provided more than one sentences.

The CESSDA "brand" and expertise

- CESSDA setting standards for collaboration of archives with journals.
- Involvement could strengthen the position of CESSDA as a consortium of respectable and reliable repositories.
- To coordinate efforts in promoting good scientific practice by publishing replication data.
- This can improve the visibility of CESSDA and SPs and their importance in the data world.
- It could help SPs for enhancing our plan of establishing a further collaboration with other journals.

Knowledge exchange, good practice and economies of scale through collaboration

- Because of the knowledge, experience and opportunity for counselling.
- Approaching journals actively is labour intensive, it would be good if CESSDA did that for all interested SPs.
- It is good to share experience among SPs, to learn from each other and to develop jointly some specific services (e.g. how to provide access for blind peer review).
- Stronger impact if we collaborate.
- CESSDA SPs can provide support for data that are truly re-usable, not just linked to a dataset without any documentation. It can also promote good practices around ethical issues in managing personal and sensitive data.
- It would increase the visibility of service providers and also simplify processes for researchers.
- CESSDA and its SPs are welcome to participate in the Research Data Journal (RDJ) and could use it "the other way around", promoting articles about data sets in partner archives.

Opportunity cost of not engaging

- Because of the importance for data transparency for publication in high quality journals.
- If CESSDA is not involved then more publishers will establish repositories of poor quality, and this might pose a threat to high-quality archives.
- This issue will be increasingly important, general support, guidance is very useful. But we should allow for different progress for archives at different stages of development.

---

5. Synthesis and recommendations

The objective of the survey and of this report was to assess CESSDA service provider capacities, policies, and services for responding to the needs of journals regarding preservation of data and replication material. The results of the survey analysis should be juxtaposed with the first report (Alvanides et al. 2021) on journal practices, requirements, and needs regarding the accessibility of data used in scientific publications. The analysis of responding SP capacities in relation to journal requirements, revealed three findings.

First, most of the archives are well represented in repository registries (although mostly with Re3ata), and fully developed in offering long-term curated data collections, quality control, personalised expert curation support and links to publications. In addition, all SPs offer currently, or will offer in the near future, training on data management planning and promotional activities and material related to data usage. These are all excellent services at the core of the SPs' mission and they can benefit journals and authors who are focusing on established social surveys that have been or can be deposited with the archives. However, publishers, editors and researchers concerned with replicable research are less likely to benefit from the current SP offerings, given the general shortage of replication support services.

Secondly, the already developed services and activities of most SPs also influence the way the SPs perceive the needs of journals and researchers who promote or require data sharing practices. It is telling that the responding SPs ranked highly those services they already offer, for example long-term curation of data collections, PID and versioning of datasets, data usage information, citations tracking and links to publications. Again, these are important services that SPs are delivering at a high standard demonstrating their awareness, but as highlighted in the first *Task 4* report (Alvanides et al. 2021), journal editors almost take these services for granted. There is scope therefore for SPs (and potentially for CESSDA's role) to promote these services to journal editors and publishers, as expectations are gradually moving further in terms of what is expected of data repositories (Sansone et al. 2020). Most of the responding SPs are aware of this shift, by identifying limitations such as handling new types of data (i.e. beyond "traditional" social surveys) or diverse datasets (e.g. resulting from inter- or multidisciplinary research), dealing with sensitive data that require controlled or restricted access and potentially size limits for larger datasets. The first *Task 4* report (Alvanides et al. 2021) highlighted that the journal editors interviewed appreciate the complexities associated with the diversity of social science data and the additional demands such complex datasets impose on sharing and replicability. Both the SPs and the editors also highlighted limitations with financial and technical resources in order to implement fully the data sharing and replicability policies envisaged. This situation leaves a void, but also an opportunity for fruitful collaboration between the interested parties.

Thirdly, there is a general positive attitude in the responses on offering data sharing and replicability services in collaboration with publishers and journals. The vast majority of

responding SPs do accept data and associated materials linked to journal articles, although only half of them have already received such deposits from journal articles. Most of those accepting data linked to journal articles do not face major technical or policy issues and have plans for expanding their services, which sets a positive precedent for the other archives. In relation to CESSDA's role, the majority of SPs responded positively and highlighted the CESSDA "brand" and expertise as driving forces for involvement. They listed various benefits of collaboration in relation to knowledge exchange, good practice and economies of scale, acknowledging the importance of expertise to be shared amongst CESSDA members. Finally, some SPs pointed out the opportunity cost resulting from lack of engagement, and the fact that the current void will be quickly filled by the journal publishers themselves, at the risk of not involving the national archives. This was also a concern raised by the editors in the first report (Alvanides et al. 2021) in relation to the understanding of policies on data protection, storage and sharing at the national and European levels.

This report has completed an assessment of the CESSDA SP's capacities, policies, and services for responding to current and potential needs of journals regarding the preservation of data (mostly) and replication material (to a lesser extent). There is a good understanding of the evolving landscape and willingness of SPs to develop additional services that fulfil the archival needs of journals, but also an expectation for CESSDA to become more involved and facilitate collaboration and sharing of expertise and good practice. The next objective of *Task 4* is to explore models for how CESSDA could coordinate SPs and to propose models for how SPs may collaborate in offering additional services to journals.

Bibliography

Alvanides, S., Štebe, J., Glavica, M., Kranjec, I., De Paula, E.M., Kleiner, B., Hegedus, P. 2021. "Assessment of Journal Requirements and Needs" CESSDA Widening Activities and Journal Outreach 2020 Task 4, DOI: 10.5281/zenodo.5554436

Crosas, M., Gautier, J., Karcher, S., Kirilova, D., Otalora, G., & Schwartz, A. (2018, March 30). Data policies of highly-ranked social science journals. https://doi.org/10.31235/osf.io/9h7ay

Resnik, D.B., Morales, M., Landrum, R., Shi, M., Minnier, J., Vasilevsky, N.A., & Champieux R.E. 2019. Effect of impact factor and discipline on journal data sharing policies, Accountability in Research, 26:3, 139-156. https://doi.org/10.1080/08989621.2019.1591277

Rousi, A.M. & Laakso, M. 2020. Journal research data sharing policies: a study of highly-cited journals in neuroscience, physics, and operations research. *Scientometrics*, 124, 131-152. https://link.springer.com/article/10.1007/s11192-020-03467-9

Sansone, S.A., McQuilton, P., Cousijn, H., Cannon, M., Chan, W.M., Callaghan, S., ... Threlfall, J. (2020, October 13). Data Repository Selection: Criteria That Matter. Zenodo. http://doi.org/10.5281/zenodo.4084763

Appendix A: email to Basecamp SP Forum

From: Brian KLEINER (Basecamp) <notifications@3.basecamp.com>
Sent: 21 August 2020 16:32
Subject: (Service Providers' Forum) CESSDA-journals outreach - survey

Dear colleague,

We are contacting you in relation to the CESSDA journals sub-project, where several CESSDA service providers are working together to assess journal needs with respect to data, as well as service provider capacities, policies, and services for responding to these needs. This sub-project is part of the larger CESSDA Widening Activities and Journal Outreach 2020: https://www.cessda.eu/About/Projects/Work-Plans/Work-Plan-2020#wid20.

Our overall aim is to formulate a coordinated CESSDA-level strategy with respect to archival services that could be provided to social science journals. To this end, we are asking CESSDA service providers to complete a brief survey to assess where we are with respect to data-related journal requirements. The responses will provide evidence for an internal CESSDA report assessing the service provider capacities, policies, and services for responding to the needs of journals regarding preservation of data and replication material. It would be very helpful to have your archive represented in our results, and we would appreciate it if you could complete the survey yourself and/or disseminate it to appropriate colleagues within your institution for completion.

The survey is estimated to take about **15 minutes:** https://limesurvey.srce.hr/281179
We would appreciate it if you could have it completed by **Thursday, September 3rd, 2020** at the latest. Please address any questions to Serafeim Alvanides at Serafeim.Alvanides@gesis.org

Thank you in anticipation,
The CESSDA-journals team
Organisation | Name
GESIS, Germany | Serafeim Alvanides (Lead)
ADP, Slovenia | Janez Štebe
FORS, Switzerland | Emilie Morgan De Paula
FORS, Switzerland | Brian Kleiner
CROSSDA (FFZG), Croatia | Marijana Glavica
CROSSDA (FFZG), Croatia | Irena Kranjec
TARKI, Hungary | Peter Hegedus

You can reply to this email or respond in Basecamp.

Appendix B: instrument for online survey

CESSDA Journals Outreach

Thank you for considering this short survey in relation to the CESSDA journals sub-project, where several CESSDA service providers are working together to assess journal needs with respect to data, as well as service provider capacities, policies, and services for responding to these needs.

This sub-project is part of the larger [CESSDA Widening Activities and Journal Outreach](). Our overall aim is to formulate a coordinated CESSDA-level strategy with respect to archival services that could be provided to social science journals. To this end, we are asking CESSDA service providers to complete a brief survey to assess where we are with respect to data-related journal requirements. The responses will provide evidence for an internal CESSDA report assessing the service provider capacities, policies, and services for responding to the needs of journals regarding preservation of data and replication material.

If you feel that you are not in a position to complete the survey yourself please forward the link https://limesurvey.srce.hr/281179 to relevant colleagues from your archive for completion. More than one person per organisation can respond to this survey. The survey is estimated to take about 15 minutes to complete and we would appreciate it if you could complete it by Thursday 23rd September 2020 at the latest.

The CESSDA-journals team

Serafeim Alvanides, GESIS, Germany
Janez Štebe, ADP, Slovenia
Emilie Morgan De Paula, FORS, Switzerland
Brian Kleiner, FORS, Switzerland
Peter Hegedus, TARKI, Hungary
Marijana Glavica, CROSSDA, Croatia
Irena Kranjec, CROSSDA, Croatia

There are 25 questions in this survey

A. Archive profile in relation to journals requirements

q0. In which country is your archive located?

Please write your answer here:

q1. Is your archive listed in any of the following repository registries in order to be findable as potential place of deposit of data related to publication?

Check all that apply

☐ Re3data
☐ FairSharing Registry
☐ Other
☐ Not listed in any registry
☐ I don't know

q1a. Would you consider being included in one of the data repository registries in the near future?

Only answer this question if the following conditions are met:
Answer was ' Not listed in any registry' at question '2 [q1]' (Is your archive listed in any of the following repository registries in order to be findable as potential place of deposit of data related to publication?)

○ Yes   ○ No

q1b. Other repository registries in which your data archive is included.

Only answer this question if the following conditions are met:
Answer was ' Other' at question '2 [q1]' (Is your archive listed in any of the following repository registries in order to be findable as potential place of deposit of data related to publication?)

Please write your answer here:

q2. Which of the following services that might be of relevance for data-related journal requirements are <u>available</u> or <u>planned in the near future</u> at your data archive?

	Yes, available	Not available, but planned	Not available, not planned	Don't know
A. Long-term curated data collections	○	○	○	○
B. Short-term self-archiving	○	○	○	○
C. Quality control, for example, data consistency or checks on data anonymisation, completeness of documentation	○	○	○	○
D. Personalised expert curation support for data and metadata preparation and deposit	○	○	○	○
E. PID and versioning for datasets	○	○	○	○
F. Relation in metadata to ORCID, Research Organisation Registry, Open Funder Registry or similar identifier registries	○	○	○	○
G. Specific data usage information or citations tracking	○	○	○	○
H. Links to publications	○	○	○	○
I. Linking to related software	○	○	○	○
J. Access to data for peer review during embargo period	○	○	○	○
K. Support for 'double-blind' peer review of data during embargo	○	○	○	○
L. Standard licences	○	○	○	○
M. Pre-registration service	○	○	○	○
N. Replication support service	○	○	○	○
O. Training, advice and active support for data management planning	○	○	○	○
P. Promotion of data usage (e.g., via blog, newsletter, trainings for future data users…)	○	○	○	○

q2a. Feel free to explain and comment in more detail about any of the services mentioned above. Also, specify any other services that you might have in your archive but were not mentioned.

Please write your answer here:

q3. Which of the services listed in question 2 do you think could be most interesting to authors as incentives to deposit their data with an archive? Please select only services that you think are important and rank them in order of importance.

Please number each box in order of preference from 1 to 16

- o A. Long-term curated data collections
- o B. Short-term self-archiving
- o C. Quality control, for example, data consistency or checks on data anonymisation, completeness of documentation
- o D. Personalised expert curation support for data and metadata preparation and deposit
- o E. PID and versioning for datasets
- o F. Relation in metadata to ORCID, Research Organisation Registry, Open Funder Registry or similar identifier registries
- o G. Specific data usage information or citations tracking
- o H. Links to publications
- o I. Linking to related software
- o J. Access to data for peer review during embargo period
- o K. Support for 'double-blind' peer review of data during embargo
- o L. Standard licences
- o M. Pre-registration service
- o N. Replication support service
- o O. Training, advice and active support for data management planning
- o P. Promotion of data usage (e.g., via blog, newsletter, trainings for future data users…)

q3a. Are there any other services, whether or not they are available in your archive, that might be of interest to authors and journals as incentives to deposit data with an archive?

Please write your answer here:

q4. Based on your archive's data collection policy and selection procedures, are there any limitations on accepting data and documentation with the following characteristics?

	No limitations	Potential limitations	Strong limitations	Don't know
A. Contain language that is not English or one of the official languages of the country you are located in?	○	○	○	○
B. Data from projects not funded by specific funding sources (e.g. private funding, non-national funding) or depositor institutional background	○	○	○	○
C. Data that require controlled or restricted access (e.g. sensitive personal data, non-anonymised data)	○	○	○	○
D. Size limit	○	○	○	○
E. Types and formats of data that are outside the scope of your data collection policy	○	○	○	○

| F. Data from other disciplines than the social sciences | ○ | ○ | ○ | ○ |
| G. Code/syntax for preparing data and/or reproducing analysis | ○ | ○ | ○ | ○ |

B. Collaboration with journals

Now we would like to ask you about any forms of collaboration that your archive might have or plan to have with journals.

q5. Does your archive accept data and associated materials linked to journal articles?

○ Yes   ○ No

q6. Do you already have data and associated materials linked to journal articles deposited in your archive?

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '10 [q5]' (Does your archive accept data and associated materials linked to journal articles?)

○ Yes   ○ No

q7. Are you facing any challenging technical or policy issues by accepting data linked to journal articles?

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '10 [q5]' (Does your archive accept data and associated materials linked to journal articles?)

○ No   ○ Yes, what kind?

q8. Would you need additional support in order to accept data linked to journal articles?

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '10 [q5]' (Does your archive accept data and associated materials linked to journal articles?)

○ No   ○ Yes, what kind?

q9. Does your archive currently have any collaboration with particular journals or publishers so that authors are recommended or obliged to deposit their data with you?

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '10 [q5]' (Does your archive accept data and associated materials linked to journal articles?)

○ Yes   ○ No

q10. Please describe this collaboration with particular journals or publishers.

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '14 [q9]' (Does your archive currently have any collaboration with particular journals or publishers so that authors are recommended or obliged to deposit their data with you?)

Please write your answer here:

q11. What kind of support do you provide to the journals or publishers that you cooperate with?

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '10 [q5]' (Does your archive accept data and associated materials linked to journal articles?) and Answer was 'Yes' at question '14 [q9]' (Does your archive currently have any collaboration with particular journals or publishers so that authors are recommended or obliged to deposit their data with you?)

Please explain why

q12. Does your archive have any plans to go in the direction of accepting data linked to journal articles?

Only answer this question if the following conditions are met:
Answer was 'No' at question '10 [q5]' (Does your archive accept data and associated materials linked to journal articles?)

    ○ Yes   ○ No

q13. What significant technical or policy changes would you have to make in order to go in the direction of accepting data linked to journal articles?

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '17 [q12]' (Does your archive have any plans to go in the direction of accepting data linked to journal articles?)

Please write your answer here:

q14. Does your archive currently have any plans for collaboration with particular journals or publishers so that authors are recommended or obliged to deposit their data with you?

Only answer this question if the following conditions are met:

Answer was 'Yes' at question '17 [q12]' (Does your archive have any plans to go in the direction of accepting data linked to journal articles?)

Answer was 'No' at question '14 [q9]' (Does your archive currently have any collaboration with particular journals or publishers so that authors are recommended or obliged to deposit their data with you?)

    ○ Yes   ○ No

q15. Do you think that a coordinated CESSDA approach to supporting social science journals and promoting CESSDA SP's would be a good way to go?

    ○ Yes   ○ No   ○ I don't know

q15a. Why do you think that this would be a good approach?

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '20 [q15]' (Do you think that a coordinated CESSDA approach to supporting social science journals and promoting CESSDA SP's would be a good way to go?)

Please write your answer here:

q15b. Why do you think that this would not be a good approach?

Only answer this question if the following conditions are met:
Answer was 'No' at question '20 [q15]' (Do you think that a coordinated CESSDA approach to supporting social science journals and promoting CESSDA SP's would be a good way to go?)

Please write your answer here:

q16. Would you be interested in talking about these issues with someone from the CESSDA-journals project?

◯ Yes   ◯ No

q16a. Please provide contact e-mail.

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '23 [q16]' (Would you be interested in talking about these issues with someone from the CESSDA-journals project?)

Please write your answer here:

q17. Do you have any relevant comments that you would like to share with us?

Please write your answer here:

Thank you!

Your survey responses have been recorded.

If you require more information about this project please contact Serafeim Alvanides, GESIS, Germany, Serafeim.Alvanides@gesis.org

Submit your survey.

Thank you for completing this survey.