Assessment of the registration of bibliographical data in the ESS Bibliography database

Professor Birger Larsen
Aalborg University
birger@hum.aau.dk

August 2018
Table of Contents

1 Introduction ............................................................................................................................................... 4
2 Methodology ............................................................................................................................................. 4
3 Background - ESS and the ESSB ................................................................................................................. 4
4 Analysis ...................................................................................................................................................... 5
4.1 Current procedures of registering data to the ESSB ......................................................................... 5
  4.1.1 Self-reporting by ESS users ........................................................................................................ 7
  4.1.2 Identification of Google Scholar publications at the University of Ljubljana .................... 11
  4.1.3 Bibliographic control ............................................................................................................... 13
4.2 Current presentation of ESS bibliographical data and proposed improvements ....................... 14
4.3 Possibilities for bibliometric analyses .............................................................................................. 24
4.4 Collaboration between NSD and University of Ljubljana and possible improvements ............ 27
5 Conclusions .............................................................................................................................................. 28
6 Acknowledgments ................................................................................................................................... 29
7 References ............................................................................................................................................... 29
8 Appendix A – ESSB data registration steps .............................................................................................. 30
9 Appendix B – Possibilities in the Google Scholar, Publish or Perish and duplicate checkers .......... 34
  9.1 Avoiding the top1000 result limit in Google Scholar by Boolean searches ......................... 34
  9.2 Using Publish or Perish (PoP) ....................................................................................................... 35
  9.3 Using reference manager duplicate checkers ............................................................................. 36
10 Appendix C – Bibliographic control ..................................................................................................... 38
Table of Figures

Figure 1. ESSB publication statistics 2000-2018 (source: http://www.europeansocialsurvey.org/bibliography/stats.htm - July 2018) ............................................................... 6
Figure 2. Very discrete notice about registering publications to the ESSB (highlight added) ................................. 8
Figure 3. Deposit requirement nearly invisible at the very end of the page (highlight added) .............................. 9
Figure 4. ESSB frontpage with Advanced search unfolded (http://www.europeansocialsurvey.org/bibliography/) ................................................................................. 18
Figure 5. Aalborg University Library (AUB) search box, interactively unfolded when clicking in the box (http://aub.aau.dk) ...................................................................................................................... 18
Figure 6. Web of Science search page .................................................................................................................. 19
Figure 7. Scopus search page. ............................................................................................................................. 19
Figure 8. ESSB result page. ............................................................................................................................... 20
Figure 9. Aalborg University Library (AUB) result page ..................................................................................... 21
Figure 10. Web of Science result page. ............................................................................................................... 22
Figure 11. Scopus result page. ............................................................................................................................ 23
Figure 12. Google Scholar search results - including number of citations in Google Scholar and the star to add a publication to ‘My Library’ (http://scholar.google.com) ........................................................................ 26
Figure 13. Scopus widget to show number of citations. Links to a list of the citing articles. (https://dev.elsevier.com/html_widgets.html) ................................................................. 26
Figure 14. Step 1 - Information about publication type, author(s), title, language and abstract ...................... 30
Figure 15. Step 2 - Information about the source, dependent on publication type. For instance for journals: journal name, ISSN, peer review status, year, volume, issue, page numbers and DOI. ................................................................. 31
Figure 16. Step 3 - Information about data use, that is ESS round/year, and which countries were used. Also contact email and any other comments. .................................................. 32
Figure 17. Step 4 - Review and confirmation of entered data - with possibility to edit the provided information ................................................................. 33
Figure 18. Venn-diagram illustrating how to split a set of between 2000 and 3000 Google Scholar results. 35
Figure 19. Publish or Perish software main window with query in progress ...................................................... 36
Figure 20. Several potential duplicate records in the ESSB and lack of bibliographic control (e.g. Breen, M vs. Breen, Michael) – highlight added. ................................................................................. 38
Figure 21. Lack of bibliographic control on publisher (Oxford Univ. vs. Oxford University) ......................... 39
Figure 22. Two different entries for the same publisher (Academia) – clicking on the publisher name in the full records does not lead to the same publisher entry. ................................................................................. 40
1 Introduction

NSD – the Norwegian Centre for Research Data – is data archive and data provider for the European Social Survey and maintains the ESS Bibliography (ESSB). Developed 10 years ago, the ESSB contains information about publications based on ESS – be it analysis of ESS data, ESS methodological research or ESS documentation. The ESSB is thus a valuable source for demonstrating the broader impact of ESS. Data for the ESSB is registered at NSD in Norway, partially using input from University of Ljubljana in Slovenia.

NSD wishes to get an assessment of the current procedures of registering data to the ESSB, including alternative methods and sources, as well as an assessment of the current presentation of ESSB bibliographical data, including possible improvement and adaptions to other similar bibliographical systems.

Specifically, this report assesses four issues:

1. The current procedure for authorization, collection and harvesting of bibliographical data, and possible alternative methods and sources for this,
2. The current presentation of ESS bibliographical data and proposed improvements,
3. Possibilities for bibliometric analyses,
4. The collaboration between NSD and University of Ljubljana and possible improvements.

The report is written from the perspective that there may be resources available to invest in improving the ESSB and its functionality.

2 Methodology

The report is based on the following sources: 1) Interviews with Knut Kalgraff Skjåk and Ole-Petter Øvrebø from the Department for Survey and Data Services, NSD and Brina Malnar from University of Ljubljana, 2) Analysis of the ESSB itself, 3) a 2017 impact study of ESS (Kolarz et al., 2017), and 4) comparative analysis of similar bibliographical systems and relevant tools.

The structured interviews were carried out online in June and July 2018 using Adobe Connect.

3 Background - ESS and the ESSB

The European Social Survey (ESS) launched its first round in 2002, following the publication of a 1999 blueprint instigated by the European Science Foundation. ESS was supported by the ESF until the end of 2012. From 2013 onwards ESS is granted status a European Research infrastructure Consortium (ERIC).

The purpose of ESS is to conduct a biennial social cross-national survey of attitudes and behaviour. The ESS uses cross-sectional, probability samples which are representative of all persons aged 15 and over resident within private households in each country. The surveys are designed with core modules (which remain the same across rounds), and rotating modules (which can change from round to round). Eight ESS rounds have

1 Available online at http://www.europeansocialsurvey.org/bibliography
been completed the last in 2016. ESS data are freely available from the ESS web site upon registration with a working email address. As demonstrated by the impact study by Kolarz et al. (2017), the ESS is a major source for social survey data with more than 100,000 registered users, and around two thirds of these having downloaded data. About one third of the registered users are non-students and 70% of those have downloaded data.

The purpose of the ESSB is defined as follows: “...to provide the ESS network, the social science community, decision makers and others with an information system that includes articles, analysis and commentaries based on ESS data (whether substantive or methodological). [ESSB hosts] a comprehensive bibliography of ESS publications and bring the network of users into contact with one another. [ESSB is] an on-line dynamic service with flexible access and retrieval, providing convenient access to all the publications within it. The system will conform to the highest bibliographical standards and thereby be a part of the international information chain on scientific publications. The system will be freely available, providing researchers, governments, organisations, decision makers and others with a unique list of social reporting on Europe.“

(from Description of Work, project ESSi, contract number RII3-CT-2006_026042 under the 6th EU framework programme)

ESS users are obliged to self-report any output they have produced based on ESS data. These self-reported publications are then included in the ESSB. However, not all users remember to do so. Kolarz et al. (2017) remark that markedly higher number of publications are registered in periods where greater efforts have been made to reminds users to do so, e.g. in 2007 and 2013. The coverage of ESSB was estimated for journal articles by Kolarz et al. (2017). They find that of 960 publications based on ESS that could be identified in the Web of Science (WoS) database up until 2016, 715 of which (76.6%) were listed in ESS. The additional publications were primarily located by searching WoS abstracts for “European Social Survey” or “ESS”. Kolarz et al. also note that the ESSB coverage of journal articles appears to have risen considerably since then based on the Google Scholar searches carried out by the University of Ljubljana (2017, p. 41). These searches are carried out annually as part of the ESS internal studies of the impact of ESS and are described in greater detail below.

4 Analysis

4.1 Current procedures of registering data to the ESSB

Basic data on the volume and growth of ESSB can be seen from Figure 1. ESSB contains at present (July 2018) a total of 3017 records for the period 2000-2018. A steady and rapid growth can be observed from 2000 to 2014 – with 452 publications as the maximum in 2014. Additional efforts were carried out in the SUSTAIN project to identify and enter missing publications in the ESSB based on the ones identified at the University of Ljubljana (Malnar, 2017c). The available funds allowed registration of primarily journal articles for the years 2012-2015. This effort is clearly visible in Table 1. The dominant publication type is journal articles (more than 51%), followed by book chapters (12%), edited volumes (10%), conference papers (8%) and reports/working papers (7%) as the major publication types.

According to data from NSD, 25% of all the publications in the ESSB as of August 2018 are additional registrations contributed by the team at University of Ljubljana. Because of the additional efforts in the

---

3 http://www.europeansocialsurvey.org/data
SUSTAIN project 74-87% of the publications identified in Google Scholar was added to the ESSB for the years 2013 and 2014 – for earlier years this share was much lower (Malnar, 2017c).

Figure 1. ESSB publication statistics 2000-2018 (source: http://www.europeansocialsurvey.org/bibliography/stats.htm - July 2018).
4.1.1 Self-reporting by ESS users

In the ESS Conditions of use\(^4\) it is requested that references to the ESS data and Data Documentation be given in particular forms as listed on the web site, and that these references are included in footnotes or in the reference section of publications. In addition, at the very end of the page it is noted that “users of ESS data are required to register bibliographic citations of all forms of publications referring to ESS data in the ESS on-line bibliography database at http://www.europeansocialsurvey.org” (see Figure 3). The latter URL is not hyperlinked. The actual registration form is accessible from the ESSB page at http://www.europeansocialsurvey.org/bibliography/ (see Figure 2) by clicking the ‘Add Publication’ button. This brings up a new page with a set of web forms to be filled in four steps (see Figure 14 to Figure 17 in Appendix A):

- **Step 1**: Information about publication type, author(s), title, language and abstract.

- **Step 2**: Information about the source, dependent on publication type. For instance for journals: journal name, ISSNs, peer review status, year, volume, issue, page numbers and DOI.

- **Step 3**: Information about data use, that is ESS round/year, and which countries were used. Also contact email and any other comments.

- **Step 4**: Review and confirmation of entered data – with possibility to edit the provided information.

Finally, a reCAPTCHA key is required to save the publication.

Although self-reporting of any publications resulting from use of ESS data is mandatory there are clear indications that many authors forget to do so or do so with a significant delay. There can be several reasons for this. First and foremost, it is not very obvious that registration is mandatory. The notice announcing this is buried deep on the ESS Conditions of use page and it is likely that many users are not aware of it at all (see Figure 3). The URL points to the main page and not the registration page - and it is not hyperlinked. The link to register publications does not make the requirement to register clear “...publications based on the ESS can be registered in the ESS Bibliography.” (my emphasis, see original in Figure 2). Also, when registering by email and when downloading data the conditions of use and the requirement to register resulting publications are not included.

\(^4\) http://www.europeansocialsurvey.org/data/conditions_of_use.html
Figure 2. Very discrete notice about registering publications to the ESSB (highlight added).
European Social Survey (ESS) - Conditions of use

Foldings. This page is intended for internal use and is not meant for public distribution. Please do not distribute this page for any use other than personal or for educational purposes.

Disclaimer. The ESS data is owned by the ESS team and the producers have no responsibility for the use of the data. All users of the data bear all risk, and any potential copyright claims are based on these risks. The ESS team, or the producers accept no liability for losses, consequential or otherwise, damages or losses arising from the use of the data, or from the interpretation, or use of the data in any way.

Figure 3. Deposit requirement nearly invisible at the very end of the page (highlight added).
Secondly, researchers have many responsibilities and registering publications may rank low in their priorities. This is certainly the experience of any University Library responsible for managing the self-registration of its university’s publications by its researchers. Registration thus needs to be as easy and non-cumbersome as possible. The data to be filled in for the ESSB are comprehensive, but also somewhat cumbersome to add.

Thirdly, it is likely that researchers have already registered their publication in their own university’s system and/or their own reference manager software, but there is no possibility to import the main bibliographical data from such systems or other sources like publisher web pages.

Fourth, the guidelines for how to cite ESS data give clear instructions for how to cite given ESS data sets. The proposed references follow bibliographical standards and look like standard bibliographical references. However, these are hard to process automatically and make precise retrieval difficult. Data citation standard and practices have developed significantly in recent years including Digital Object Identifiers (DOIs) for data sets which the ESS and ESB could benefit from.

RECOMMENDATIONS

R1. **To make it much more clear to registered users the registration of publications is mandatory, and to formulate and carry out a strategy for building a culture among ESS users where they remember to register their publications to ESSB to a much greater extent.** This includes making the requirement much more prominent on the ESS web site including with clearly visible, direct links to the registration page, and should include these also when registering for download and at each download, e.g. by including the requirement and direct registration link in the emails and in the data files themselves. It should also be considered to send emails to relevant registered users on a regular basis, e.g. once or twice per year, to remind them to register any new publications. Inspiration to a greater involvement of ESS users can be gained from social networking sites for scientists such as ResearchGate, Academia.edu or Mendeley. These sites now regularly send personalised emails to registered users seeking to involve them actively in several ways. This includes notifying users of new publications that the site has discovered that may be authored by the user and asking to confirm the correctness of the data, and various status updates on the number of publications, reads, citations etc of that user’s publications. By carefully timed emails with content of interest to the ESS users a more active relationship can be built and many more chances to remind authors to register any additional publications. These emails should probably be limited to the user types who are likely to publish, i.e. excluding student downloaders, and an opt-out link be included in every email. The **frontpage of the main ESS web site can also be used to highlight recent high impact publications** (e.g. from the last 5 years) in prestigious publication channels and/or with high citation numbers – with encouragement to register your own publication. And the **ESSB can be made more prominent on the ESS frontpage** – as it is now it is hidden among many other links at the very bottom of the page.

R2. **To review the recent developments in data citation standards and practises and to analyse how the ESS and ESB could benefit from these.** For example, assigning DOIs to ESS data would in time greatly increase the possibility of identifying ESS data sets and retrieving publications that cite them automatically. For this DOIs need to be assigned at NSD at **appropriate levels of granularity**, and should be included in data set downloads. The level of granularity could be the

---

5 See recommendation R33 for how to include citations.
ESS rounds, but it should be analysed if finer granularity would be beneficial. In addition, **authors need to be encouraged to use these DOIs when citing ESS data** in publications. This culture change will take time, but the effort is likely to be worthwhile. Experiences with this process and other data citation issues are accumulating, e.g. at the Australian National Data Service⁶ which provide detailed information and inspiration. If a data citation culture can be built over time where authors use ESS specific DOIs automatic identification of publications that use ESS data for the ESSB has a much greater chance of being automated.

**R3. To make it easier to import bibliographical data from other sources.** This includes structured import from reference managers (e.g. RefWorks, EndNote, Reference Manager, Zotero, Mendeley and similar systems⁷) and university Research Information Management systems such as Elsevier’s PURE, making it possible to paste a DOI and retrieving the bibliographical data automatically⁸, and to import from other databases like Scopus, Google Scholar etc. For each system it is necessary to configure an importer so a **survey of the most popular systems** and web sites for reference management among ESS users will be a good idea to help prioritise. Not all data needed for ESSB registration can be imported this way – but there is a good chance that the data types from steps 1 and 2 can be imported (see Figure 14 and Figure 15).

4.1.2 **Identification of Google Scholar publications at the University of Ljubljana**

As part of the management of the ESS ERIC, Dr. Brina Malnar at the University of Ljubljana carries out comprehensive searches in Google Scholar annually. The purpose of this work is to identify and analyse publications that use ESS data and thereby to gain insight into the impact of ESS. The identified publications are forwarded to NSD, where they are checked and verified, and if within the inclusion criteria, added to the ESSB.

The searches are carried out in Google Scholar to maximise the chance of obtaining a comprehensive coverage. The searches are carried out through the Google Scholar web interface, with the language set to English. The search phrases used are “European Social Survey” and “ESS” and the search is limited to the year under analysis. It is the experience of Dr. Malnar that reference practises vary as to how the ESS is cited: about 40% cite the ESS website, about 20% refer to the used round explicitly and there is great variety across academic fields. The Google Scholar results are then copied into a table in MS Word. By manipulating the settings 20 records can be shown per result page, and Google displays no more than the first 1000 results – regardless of how many hits there are.

Once downloaded the records are further processed. As there are many duplicates in Google Scholar a major task is to identify, merge or remove these. The remaining records are then cleaned and coded for a large number of variables listed in Table 1.

---


⁷ Inspiration for such importers can be found for instance in these examples:

- [https://www.refworks.com/refworks/help/Overview_Building.htm](https://www.refworks.com/refworks/help/Overview_BUILDING.htm)
- [https://canterbury.libguides.com/endnote/extra-importing](https://canterbury.libguides.com/endnote/extra-importing)
- [https://guides.lib.unc.edu/Covidence/move](https://guides.lib.unc.edu/Covidence/move)
- [https://www.zotero.org/support/kb/importing_records_from_endnote](https://www.zotero.org/support/kb/importing_records_from_endnote)

⁸ See [https://www.crossref.org/labs/citation-formatting-service/](https://www.crossref.org/labs/citation-formatting-service/)
The data is used for ESS internal studies of the impact of ESS, resulting in reports like the *ESS Annual Bibliometric Reports* (e.g. Malnar, 2017a) and the *ESS Item Usage Reports* (e.g. Malnar, [2017b]) – some of which are deliverables in the ESS ERIC. Publication data is also forwarded to NSD for verification and possible inclusion in the ESSB.

Table 1. Coded variables on publications based on ESS data done at the University of Ljubljana.

| • Basic bibliographic data |
| • Publication type |
| • Open access status |
| • Language |
| • Publication year |
| • Number of Authors |
| • First author, including country and affiliation |
| • Topic(s) |
| • Sub-field (journals only) |
| • ESS round(s) data used + reason if multiple used |
| • Countries used |
| • Level - micro/macro study |
| • Other sources used |
| • NUTS region |
| • Explicit data weights |
| • Response rates |
| • Links to ESS webpages |
| • Policy references (policy keywords + domain) |
| • Number of items + which items used (items being ESS survey questions) |
| • Modules used |

The work to identify publications using ESS data done at University of Ljubljana represents a significant and impressive effort. The work is largely manual in nature and involves a lot of results clicking and copying in Google Scholar and manual removal of duplicates as well as skimming through the full-text of publications to accomplish the detailed coding.

Google Scholar does not provide an API that allows bulk download, and is notorious for containing many duplicates. This makes it pertinent to consider alternative data sources that have such features. More professional systems include the commercial Web of Science and Scopus citation databases. Both index primarily scientific and scholarly journals, WoS some 12,000+ journals and Scopus some 22,000+ journals, and both continuously register citations to these publications making citation analysis possible. As discussed in Section 3, Kolarz et al. (2017) found that a similar number of journal articles could be identified in Web of Science and in the recent years of ESSB based on self-reporting and Google Scholar searches. No comparison appears to have been made on an article-to-article basis, and it is not clear how large a proportion of the articles are the same in both sources – only that they number approximately the same. Web of Science thus is a possible alternative to Google Scholar for journal articles. Scopus is known to index a larger number of journals than Web of Science including European ones (Mongeon and Paul-Hus, 2015), and thus has the potential to cover an ever larger proportion of ESSB relevant journal articles. The main weakness of both databases is however that they mainly cover scholarly journal articles. Both index a limited number of books and conference proceedings, but with no comprehensive coverage. However, as
can be seen from Figure 1, almost 40-50% of the publications in a given year of ESSB are non-journal articles. It is not known how many of these have been self-reported, but changing to Web of Science or Scopus there is a risk that a large proportion of these non-journal articles would be lost from the ESSB. In particular, the registration of books, book chapters and edited volumes would suffer. This is serious as they are the three major publication types following journal articles. Reports and working papers, which are important indicators of wider societal impact, would also be lost. In addition, Dr. Malnar relies on the full-text capabilities of Google Scholar to carry out the detailed coding of e.g. ESS round, countries, ESS items and policy references.

An alternative to the web interface of Google Scholar is the Publish or Perish (PoP) program. Created by Professor Anne-Wil Harzing, PoP facilitates the retrieval and analysis of citation data from Google Scholar and other sources. PoP has the same limitation in that it can only retrieve 1000 records in a single search, but all of these can be exported as a single file – reducing the need for much manual clicking in the initial download phase. A new challenge reported by Dr. Malnar is that the number of hits in recent years for a given year of ESS publications now exceed 1000 hits before the removal of duplicates. This means that only the first 1000 documents presented by Google can be retrieved. As the ranking criteria of Google Scholar are only partially known this introduces uncertainty in which publications are included in the download and which are excluded. Appendix B presents ideas for avoiding this using Boolean operations.

RECOMMENDATIONS

R4. To continue to use Google Scholar as the main source of data for the identification of ESS publications for impact assessment at University of Ljubljana – despite its obvious shortcomings. The main reason for this is the risk of losing non-journal publication types from the ESSB.

R5. To analyse and review the current steps in using Google Scholar and to hire student assistants to carry out some of the simpler and more time-consuming tasks. This includes exploring the use of the Publish or Perish program in the pipeline and exploring further the capabilities of the Google Scholar web interface. Automatic duplications checkers such as found in references managers may also aid in identifying the majority of these automatically. Ideas for the issues above are discussed in Section 4.3 and Appendix B.

4.1.3 Bibliographic control

A few cases of potential challenges with bibliographic control came up while interacting with the ESSB. Appendix C documents these. They include what appears to several entries for the same publication (Figure 20), different versions of the same author name ("Breem, M", Figure 20), different versions of the same publisher name, and different entries for the same publisher (Figure 22). Several of these can be the result of inaccurate self-reporting, or errors in imported Google Scholar data. With self-reporting playing such a prominent role in the ESSB identification of publications there is an increased risk of duplicates as authors may not be aware of bibliographic standards and rules – with increased need for bibliographic control.

R6. To analyse the current level of bibliographical control, including identification of and correction of errors to understand their causes and to improve bibliographical control in the ESSB.

R7. In case of name variants (e.g. the same author writing his/her name in different ways in different publication), to ensure that all name variants can be accessed through any of them.

---

9 See Harzing (2007) and https://harzing.com/resources/publish-or-perish
4.2 Current presentation of ESS bibliographical data and proposed improvements

Based on the statement in Section 3, the main purpose of the ESSB can be summarised as: hosting a freely available comprehensive bibliography of ESS publications conforming to the highest bibliographical standards, with flexible access and retrieval freely available, providing researchers, governments, organisations, decision makers and others with a unique list of social reporting on Europe – and to bring the network of users into contact with one another. This is however a very broad definition and it does not make it very clear who, outside the ESS, would search the ESSB and with what purpose. What different and distinct needs would the bibliography serve for various target groups, e.g. researchers, students, journalists, policy makers? What needs do current users of the ESSB have? What additional needs could the ESSB serve for current non-users, and why do they not use the ESSB? A survey and analysis of current users and potential uses could help set up a strategy for this and to prioritise efforts to improve the ESSB.

This section analyses both the bibliographical data and the ESSB interface. The data collected for registration in the ESSB can be seen from Figure 14 to Figure 17. They include standard bibliographical data as well as data on which ESS rounds and countries were used. This data is saved in a relational database and indexed in an Apache Lucene search engine for retrieval and display on the ESSB web page. According to Google Analytics data provided by NSD, in the period July 2017 to July 2018 the ESSB pages had more than 30,000 pageviews, 20,000 of these unique pageviews. About 30% of these were to the front and search pages – the rest presumably go to individual publications, 4 of which were in the top10 pages visited.

The front page of the ESSB is a search page with an image, a header and brief information about the ESSB content. It has a search box and button and links to 1) unfold an advanced search page, 2) a pop-up page with search tips, and 3) a page with the complete bibliography listed. On the right-hand side is information and link to register new publications as well as a linked list of abbreviated titles of the 10 most recent publications added to the ESSB. The page has a clean and simple layout and it not overloaded. It does not however give much instruction on how to use it, and it looks somewhat outdated in terms of design.

The search result page of the ESSB contains a table with up to 20 search results, and links to additional result pages at the top and bottom in case of more than 20 hits. The number of hits per result page is fixed to 20 and cannot be changed. The number of hits is indicated at the top left. The results are in tabular format with gridlines visible. Headers are in bold and include: Title, Authors(s), Year, Type and Published in/by, and the result can be sorted by clicking the headers, but there is no indication of which column that currently determines the sorting. The default sorting appears to be descending by publication year. Publication titles are hyperlinked and when clicked leads to a page with that record. There is no visible link back to the ESSB search page in case the query needs to be reformulated. The font size of the whole page is quite small as a lot of information is crammed in. Especially the number of search results is tiny and without space around it can easily be overlooked.

A single ESS publication record page contains the bibliographical data of the publication in tabular format. Selected elements are hyperlinked with different functionality: DOIs will lead a DOI resolver (that typically links to a full-text version at a publisher), author names will execute a search for that name in ESSB, as will

---

10 See http://www.europeansocialsurvey.org/bibliography/ or Figure 2.
11 See Figure 8.
journal names, anthology titles, and publisher names. In case of anthologies (i.e. edited volumes), if there are separate records for each contribution in ESSB with a link to the anthology, these are shown and linked under the anthology. At the top there is a ‘Return to search results’ link, and at the bottom a ‘Google this publication’ link. The latter executes a Google Web search with the publication title words and author last names, which opens in a separate window. On the top-right information about the ESS data used is given in a single list; ESS rounds first, then ESS countries.

RECOMMENDATIONS

R8. To survey and analyse the needs and wishes of current users and potential uses to help set up a strategy for prioritising future efforts to improve the ESSB. Existing and active users can be identified through their registrations and download activity. Potential users that currently do not use the ESSB, but that could benefit from it are harder to identify directly – some creative thinking needs to go into identifying these among the stakeholders.

R9. To carry out recurrent log analysis of ESSB web site usage. This includes Google Analytics to understand the web site traffic, e.g. where the users come from, where they enter and leave the site, bounce rates (single page visits only), etc. The search logs of the ESSB search page, search results, and single records pages should be analysed. In particular, 0-hit searches where the user finds no results should be analysed regularly to gain understanding of where and why users do not find what they look for.

ESSB front page (search page):

R10. The search page should indicate what can be searched for. It is not immediately obvious if you can search for title words, authors, publishers, years, publication type etc. Some of this information is given in the Search Tips, but few will look there or in the Advanced search as a first thing. AUB (Figure 5) automatically displays information about the collection. Web of Science (Figure 6) and Scopus (Figure 7) has field selectors indicating what type data is searched. Examples of possible queries could be given.

R11. The search box and search button are relatively small and not prominent enough. AUB (Figure 5) has a very large search box, Web of Science (Figure 6) highlights the box and uses clear contrast on the search button, and Scopus (Figure 7) uses an airy layout to bring focus to the search box, and a symbol and contrast on the search button.

R12. The cursor should be ready to type in the search box when loading the page (i.e. set focus in the search box). Assuming the ESSB users would want to search they should not have to click in the search box, but should be able to type directly.

R13. In advanced search several items can been searched simultaneously (e.g. from a PC by ctrl-clicking several document types) but this is not documented.

R14. Consider making additional fields searchable/limitable. The searchable fields appear to be: title, abstract, author – with limits to author country, document type and publication year in Advanced search. Institution/affiliation, language, source (e.g. journal name or publisher) as well as ESS rounds and ESS country could be made searchable or limitable – depending on user needs and scenarios.

R15. The image on the ESSB page is small and low resolution. It probably represents ESSB publications, and while providing some air on the page, it serves no real function and can be removed.

R16. The spacing between elements becomes cramped when unfolding Advanced search (Figure 4) – the heading over the lists are too close to the elements above.
R17. The ESSB page is static and does not seem alive – it does not use modern web design principles. AUB (Figure 5) unfolds extra information when the search box is clicked, Web of Science (Figure 6) use highlights and Scopus (Figure 7) changes elements to a contrast colour on mouseover.

R18. It is hard to guess what is behind the link to the complete bibliography. It can easily be misread to state that the bibliography is formatted in Harward format, without realising that the complete bibliography can be seen. Using action words a la “Click here to see the complete bibliography (xxxx publications)” would make the link much clearer.

R19. The whole web site and the ESSB are not well-suited for mobile browsers. It should be changed to a responsive web page design that adapts to the type of device.

ESSB search result page:

R20. To make it possible to show more than 20 results per result page, e.g. 50, 100, 200 all. In academic search systems users are often willing to view more hits than in web search engines. Numbering the results can also be an aid in keeping an overview regardless of number of results shown.

R21. To indicate which column the results currently are sorted by and if this is ascending or descending. In other systems this is typically done with little highlighted arrows.

R22. Add faceted navigation to the result page. This utilises the existing ESSB data to create additional navigation and findability possibilities with a given result set. AUB (Figure 9), Web of Science (Figure 10) and Scopus (Figure 11) all have facets on the left hand side of their interfaces. The facets are built by extracting information from the bibliographical data fields of the records in the current search, and the top items from each field are presented as hyperlinks along with the number of records that will result from clicking a link. Such facetted navigation allows users to quickly drill down on all of the bibliographic data and reduces the search to a more manageable size. In principle almost all bibliographical fields can be displayed as facets. The facets can be a fixed list in a fixed order, or be dynamic depending on the publication type or the frequency of the facet contents. See Russell-Rose and Tate (2013) for inspiration (especially Chapters 6 and 7).

R23. The explicitly tabular format for search results is rarely seen these days, but it does give a good overview of a lot of information and obvious sorting possibilities. A similar tabular format, but more airy and modern can be seen in Scopus (Figure 11), and may serve as inspiration. The ‘View abstract’ dropdown may be useful also for ESSB, and the “Google this publication” link could be placed next to it along an updated citation count (see recommendation R33 below).

R24. Add a link to get back to the search page.

R25. Increase the font size generally in particularly on the number of results – and add space around this line.

ESSB single record page

R26. Overall the layout gives a good overview of the bibliographical information. There should be more space around the top and bottom links to make them stand out more. The font size of most labels is very small and should be enlarged.

R27. Links to external sites should open in a separate window. The ‘Google this publication’ does so and the DOI resolver should do the same.

R28. Linking elements like authors and publishers is a good idea – it makes it easy to browse through the ESSB. But when there is only one instance in the ESSB it may create a false expectation that because there is a link there will also be additional information behind it. Elements where there are no additional records besides the current one could be exempt from the hyperlinking.
Alternatively, the number of records behind the link could be displayed, e.g. in a parenthesis or on mouseover.

R29. The ‘Google this publication’ link is a good idea as many publications are available on the open web or have a publisher page associated to it. It can be considered to direct this search to Google Scholar instead of Google Web search as less noise will result from limiting the search to academic publications. On the other hand, if the publication cannot be found in Google Scholar, web search may be preferable. Both could be offered “Search this publication in: Google – Google Scholar”.

ESSB Publication Statistics page

R30. The numbers in each cell in this page could be hyperlinked to a search, so that a list of those publications appears with a single click. This would offer an alternative access to ESSB.

---

12 See http://www.europeansocialsurvey.org/bibliography/stats.html
ESS Bibliography

The ESS Bibliography contains information about publications based on the cross-national European Social Survey. This includes analyses of ESS data, ESS methodology research and descriptions and documentation of the ESS.

Figure 4. ESSB frontpage with Advanced search unfolded ([http://www.europeansocialsurvey.org/bibliography/](http://www.europeansocialsurvey.org/bibliography/))

Figure 5. Aalborg University Library (AUB) search box, interactively unfolded when clicking in the box ([http://aub.aau.dk](http://aub.aau.dk))
Figure 6. Web of Science search page

Figure 7. Scopus search page.
Figure 8. ESSB result page.
Figure 9. Aalborg University Library (AUB) result page.
### Results: 383

You searched for: **TOPIC**: Larsen B...More

#### Filter results by:
- Open Access (120)

#### Publication Years
- 2018 (2)
- 2017 (2)
- 2016 (2)
- 2014 (2)

#### Web of Science Categories
- GEO SCIENCES MULTI-DISCIPLINARY (17)
- GEOGRAPHY PHYSICAL (14)
- RHEUMATOLOGY (12)
- CHEMISTRY PHYSICAL (12)
- GEOCHEMISTRY GEOPHYSICS (12)

#### Document Types
- ARTICLE (55)
- PROCEEDINGS PAPER (54)
- BOOK REVIEW (24)
- REVIEW (23)
- BIOGRAPHICAL ITEM (3)

#### Organizations Enhanced
- UNIVERSITY OF CALIFORNIA SYSTEM (5)
- NERC NATIONAL ENVIRONMENT RESEARCH COUNCIL (3)
- NERC BRITISH ANTARCTIC SURVEY (3)
- HELWIGS ASSOCIATION (3)
- ALFRED WEGENER INSTITUTE HELWIGS CENTER FOR POLAR

#### Search Results:

<table>
<thead>
<tr>
<th>#</th>
<th>Title</th>
<th>Authors</th>
<th>Journal/Conference</th>
<th>Year</th>
<th>Pages</th>
<th>Type</th>
<th>View Abstract</th>
<th>Full Text from Publisher</th>
<th>Usage Count</th>
<th>Times Cited</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Antarctic ice shelf disintegration triggered by sea ice loss and ocean swell</td>
<td>Massom, Robert A; Scambos, Theodore A; Bennetto, Luke G; et al.</td>
<td>NATURE</td>
<td>2018</td>
<td>150, 770, 383 – 386</td>
<td>Article</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Late Quaternary deglacial history across the Larsen B embayment, Antarctica</td>
<td>Jeong, Ara; Lee, Ja; Seong, Young-Hee; et al.</td>
<td>QUATERNARY SCIENCE REVIEWS</td>
<td>2018</td>
<td>109, 144–148</td>
<td>Article</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Development of a qPCR assay to detect and quantify ichthyotoxic flagellates along the Norwegian coast, and the first Norwegian record of Fibrocapsa japonica (Rhaphidophyceae)</td>
<td>Engvold, Anette; Strand, David Gran; Skadi Haavind, Sandra, et al.</td>
<td>HAMMAR ALGAE</td>
<td>2018</td>
<td>75, 105–117</td>
<td>Article</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Antarctic ice shelf near-surface snowmelt detection based on the synergy of SSM/I data and QuikSCAT data</td>
<td>Wang, Xing-Dong; Li, Xin-Li; Wang, Cheng; et al.</td>
<td>GEOSCIENCE FRONTIERS</td>
<td>2018</td>
<td>9, 3, p. 553–563</td>
<td>Article</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Changing pattern of ice flow and mass balance for glaciers discharging into the Larsen A and B embayments, Antarctic Peninsula, 2011 to 2016</td>
<td>Rott, Helene; Jabot, Wael Abdel; Walter, Jan, et al.</td>
<td>CRYOSPHERE</td>
<td>2018</td>
<td>12, 4, p. 1273–1291</td>
<td>Article</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>On optimal investment with processes of long or negative memory</td>
<td>Chiu, Hui N; Basravi, Wilkins</td>
<td>STOCHASTIC PROCESSES AND THEIR APPLICATIONS</td>
<td>2018</td>
<td>128, 4, p. 1095–1113</td>
<td>Article</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Increased arterial stiffness in rheumatoid arthritis and its relation to disease activity: A cross sectional study</td>
<td>Younsel, Ghada; Ameer, Nadhiah Taher; Gaber, Wafaa, et al.</td>
<td>EGYPTIAN HEART JOURNAL</td>
<td>2018</td>
<td>70, 1, p. 35–40</td>
<td>Article</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>A Mathematical Model of Melt Lake Development on an Ice Shelf</td>
<td>Buzzard, S. C.; Fethoun, D. L.; Fieno, D.</td>
<td>JOURNAL OF ADVANCES IN MODELLING EARTH SYSTEMS</td>
<td>2018</td>
<td>10, 2, p. 262–263</td>
<td>Article</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Comprehensive cluster-theory analysis of the magnetic structures and excitations in CoC2 crystalline center</td>
<td>Jeon, Jinseuk; Lee, Joonseok; Jeon, Kwonho; et al.</td>
<td></td>
<td></td>
<td></td>
<td>Article</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 10.** Web of Science result page.
Figure 11. Scopus result page.
### 4.3 Possibilities for bibliometric analyses

The popularity of metrics of downloads, use, citations is indisputably rising, and it is well-worth to consider to what extent and how this type of data and analysis can be integrated into ESSB – especially because the ESSB is made up from publications that cite ESS data. Bibliometric analysis could serve two purposes. First, data on ESS data sets usage can aid the ESS ERIC Core Scientific Team (CST) in their management of ESS, and secondly appropriately presented bibliometric data has the potential to engage ESS users more with ESSB.

In brief, bibliometric analyses include several different types:

- **Publication analysis** – quantitative studies of research publications, e.g. the publication activity in a given country or region, speciality or topic – often carried out over several time periods
- **Citation analysis** – quantitative studies of the citations received by a set of publications
  - Can be analysis of the citing publications – e.g. which countries or institutions do they come from, or which journals or topics?
  - Can be quantitative analysis of the amount and nature of citations – e.g. the average number of citations per publication, h-indexes, field-normalised indicators etc.
  - Can include network visualisation, e.g. co-author networks, journal networks etc.

A self-evident point nonetheless worth noting is that citations evolve (grow) over time. This means that citation-based bibliometric indicators may need to be recalculated periodically.

The *ESS Annual Bibliometric Reports* (e.g. Malnar, 2017a) and the *ESS Item Usage Reports* (e.g. Malnar, [2017b]) are both prime examples of very thorough analysis of citing publications and analysis of the amount and nature of citations. The 2017 ESS Annual Bibliometric Report has e.g. basic analysis of the citing publications in section 1 (Mapping research topics) and Section 2 (The geography of ESS Authors). The remainder of the report contains detailed analysis of the use of ESS data, including national data, core and rotating modules, ESS items etc. The analysis is made possible, first by the identification of citing publications and secondly by the detailed coding of these. Both involve considerable efforts, and it is my assessment that it would be almost impossible to automate the identification of publications and extensive coding of data.

The ESSB Publication Statistics is another example of a (simple) analysis of citing publications. The table gives information about the growth of ESSB over time and across publication types, and by using the dropdown menu the productivity of different countries can be studied.

The basic object of study in both are the ESS data sets – that is the ESS rounds, countries and items is the starting point for the analysis, and in a sense, this can be used to study which of the ESS data sets and subsets are most popular and which are underused – as done in the *ESS Annual Bibliometric Reports*. A different perspective is obtained by changing the starting point of the citation analysis from the ESS data sets to the citing publications themselves. By doing so we would learn about the impact not of the data sets directly, but the impact of the publications that are based on ESS data. These publications of course can themselves be cited, and by studying their impact we get an indication of how well these publications are received, and by extension a second-order impression of the impact of ESS. It is this kind of analysis that is included in the ESS impact study by Kolarz et al. (2017) represented by the bibliometric analysis carried out by CWTS (section 6.2.1 in that report). CWTS studied the academic impact of the publications citing ESS data, and were able to analyse e.g. the amount of citations received, average citation rates, relative

---

13 See Figure 1 and [http://www.europeansocialsurvey.org/bibliography/stats.html](http://www.europeansocialsurvey.org/bibliography/stats.html)
(normalised) citation rates, publications in top% percent of their fields, collaboration rates (co-authorship) nationally and internationally etc.

The results of the CWTS citation analysis are very interesting, finding for instance very high impact and significantly more top 10% cited publications in the respective fields and universities among publications in the ESSB compared to non-ESS-based publications, as well as pinpointing individual top-cited publications that has received many more citations than other publications in their fields. CWTS is one of the few centres that can reliably normalise raw citation counts and compare them to other publications in relevant fields at different levels of granularity. This allows CWTS to estimate for instance normalised citation rates that take into account the citations rates of given publication types at a given time period in relation to the same field. This is necessary because different fields are known to display difference citation rates, and direct comparison of citation rates of publications from different fields can lead to misinterpretations.

This expressivity and flexibility of analysis comes at a cost though: it is only made possible because CWTS subscribes to the full Web of Science source data, and because CWTS uses its considerable and longstanding research experience and infrastructure to clean and prepare the data for analysis. CWTS are world leaders in this type of analysis, but the cost of data and maintaining the infrastructure typically means that such fine-grained and extensive analysis is somewhat expensive. The analysis can however be carried out relatively quickly once the basic set publications for analysis have been identified. Note that CWTS can primarily analyse journal articles because these are the main unit in Web of Science.

Less sophisticated citation data can however be collected, and simple analysis can be carried out. At the time of collecting Google Scholar data at University of Ljubljana, the number of citations could be noted. The current number of citations is given in the web interface of Google Scholar (see Figure 12), but is not saved at present. As the data collection at University of Ljubljana occurs no more than 1-2 years after publication only the early citations can be discovered in this way. If the star is clicked on a publication, it will be added to ‘My library’ in Google Scholar. This is a list of publications saved in Google Scholar to a particular profile. Items on the list can be labelled (e.g. “ESS 2016 publications”) for easy retrieval later on, including updated citation counts. If every single Google Scholar hit is added to the list, citations to the publications of a given year can then be retrieved from Google Scholar at a later date. Note that the same restriction on the number of publications for display also applies to ‘My library’ – only 20 at the time can be displayed. This solution to obtaining citation counts and later updated counts thus involves a lot of manual labour. Note that is seems that the 1000 document limit also applies to a given label in ‘My library’ (this is as so many things in Google Scholar undocumented).

APIs that allow programmable access to their data are offered both by Web of Science and Scopus. Where the Web of Science API requires a subscription, the Scopus API is offered free of charge. This offers a range of interesting possibilities. Included in the API is a service that for a given Scopus document ID will retrieve the current number of citations of that publication including a label that can be displayed on a web site (see Figure 13). This label could be added e.g. on in the ESSB search result page, or in ESS publication

---

14 For instance, the Mean Normalised Citation Rate (MNCR) could be 1.6 citations per publication in the micro-subfield of “European parliament, European commission, Lisbon, euroscepticism” and 3.2 in the “corruption, financial development, privatization, bribery, economic freedom” subfield – meaning that a publication in the first field with 2 citations does better than expected, whereas a publication with 2 citations in the latter field does worse than expected(!). Example data from Kolarz et al. (2017, p. 52, Table 9).

15 See http://wokinfo.com/products_tools/products/related/webservices/

16 See https://dev.elsevier.com

17 See examples at https://dev.elsevier.com/html_widgets.html
record pages. For this to work the Scopus document ID needs to be known. Another part of the API, the Scopus Search API\(^\text{18}\) facilitates searches using the Scopus query language\(^\text{19}\). A search combining data from the ESSB can be executed (for instance on title terms and the first author last name), and the Scopus ID retrieved. Thus, it should be possible automatically to add ‘Cited x times in Scopus’ labels in ESSB using these two steps\(^\text{20}\).

---

**RECOMMENDATIONS**

**R31.** To continue and extend the support for the work done at University of Ljubljana by Dr. Malnar to identify and analyse publications that use ESS data. The work is of very high quality, the results interesting and vital to ESS and very hard to automate – especially given that the volume is in hundreds and not tens of thousands each year. Student help could alleviate some of the work load and free resources to further develop the data collection and analysis.

**R32.** If sophisticated citation analysis of journal articles in the ESSB is desired, contract this from CWTS. CWTS has the data sources, infrastructure and experience to interpret the results and to carry this out reliably, including proper normalisation techniques. Adding Web of Science ID numbers to the records in the ESSB will aid in this process and make a direct match to ESSB records possible.

**R33.** Investigate the possibility offered by the Scopus APIs for retrieving up to date citation counts from Scopus automatically. This seems to be the most promising solution for automatically adding citation counts to ESSB records and appear to be subscription free.

**R34.** If student help can be hired, carry out the manual work needed, investigate the possibilities offered by the Google Scholar ‘My library’ function for collecting up to date citation data from

---

\(^{18}\) See [https://dev.elsevier.com/search.html#!/Scopus_Search](https://dev.elsevier.com/search.html#!/Scopus_Search)

\(^{19}\) See [https://dev.elsevier.com/tips/ScopusSearchTips.htm](https://dev.elsevier.com/tips/ScopusSearchTips.htm)

\(^{20}\) See The Abstract Citation Counts API may be able to achieve this in a single step: [https://dev.elsevier.com/scopus.html#!/Abstract_Citation_Count/AbstractCitationCount](https://dev.elsevier.com/scopus.html#!/Abstract_Citation_Count/AbstractCitationCount)
Google Scholar, e.g. annually. The citation counts can be displayed in the ESSB search result page, or ESS publication record pages.

4.4 Collaboration between NSD and University of Ljubljana and possible improvements

The comprehensive coverage of the ESSB exists because of the symbiosis between NSD and the work carried out by Dr. Malnar at the University of Ljubljana. NSD runs the ESSB and receives self-reported publication and verifies these as well as the publications identified by Dr. Malnar’s comprehensive Google Scholar searches. The work at the two sites is however carried out in two separate processes and with different purposes and it is worth considering how both can better support each other, and what new synergies can arise.

Comparing the number of publications reported in the ESSB and in Dr. Malnar’s reports it seems that there is not a 100% overlap between the two sources. This may be due to different inclusion criteria, but it highlights the desirability to be able to directly compare the two sources and more systematically identify and analyse any differences in coverage. At present this is not possible because there is no direct link between the two. The potential gains in comprehensiveness of the ESSB coverage are large – as demonstrated by the SUSTAIN project (Malnar, 2017c).

RECOMMENDATIONS

R35. To investigate the possibility of directly comparing and linking the ESSB and Dr. Malnar’s data, e.g. through ensuring that the record IDs of one source is included in the other. At NSD this should not be a large effort – all Dr. Malnar’s publications are verified against ESSB records and Dr. Malnar’s IDs can be added to the ESSB. This offers new possibilities including:

- Identifying missing self-reported publications in Dr. Malnar’s data and adding these to the analyses of ESS impact.
- Importing the coded data from University of Ljubljana into the ESSB as an addition to existing data. The data is high quality and very detailed on a large number of aspects of ESS data use. These additional data can be displayed in ESSB single record pages, and some in the search result page. In addition, new search or limit possibilities arise, e.g. limit to open access publications, search/limit by topic, sub-field/domain, limit to level of study (micro/macroe), listing of other sources used, search/limit to policy references, search/limit to modules and ESS items used etc.
- Including a link to the full-text of a publication from Google Scholar. Google Scholar links in most cases to a full-text version of a given publication, typically on the publisher’s web site links. These links are collected as part of Dr. Malnar’s workflow and could be added to the ESSB records.
- Adding the country of all authors’ institutional affiliation to Dr Malnar’s data and analysis (for resource reasons only the first author’s country is coded at present) – giving a more full picture ESS data use across countries, as well as enabling analysis of collaboration between countries (see examples of collaboration analysis in Kolarz et el., 2017)

R36. To reconsider the policy of limiting the data collection from Google Scholar to publications in English only. At present only English language publication are collected. There seems to be two reasons for this: without the limitation there are far more than 1000 hits in Google Scholar for a single year, and the detailed coding cannot be carried out without proficiency in the language of

21 Appendix 2 of the 2017 ESS Annual Bibliometric Report (Malnar, 2017a) lists 3554 publications for the period 2001-2016, but the ESSB contains at present no more than 3017 publications.
the publication. The limitation to English only publications does however bias the analysis of ESS and also to some extent the content of the ESSB (as a significant 25% of the records in the ESSB are identified this way). If the ESS CST wishes to extend the analysis to all languages more resources should be allocated to achieve this.

R37. Do a **comparative detail analysis of the data fields registered at both sites** to ensure consistency, and that each registers what the other needs if possible given resources. For instance, abstracts are not stored separately in the pipeline at University of Ljubljana. But NSD needs the abstracts for their records, and for the majority this can be done with little effort by the team at University of Ljubljana (the major exception are book chapters and some conference papers).

5 Conclusions

Through the efforts of NSD and University of Ljubljana the ESSB provides, as far as can be assessed, a quite comprehensive coverage of in particular English language publications that are based on data from ESS. Through a basic search functionality, the ESSB fulfils its main purpose by offering flexible access and retrieval interface to these publications, and no severe problems have been identified in relation to this.

The analysis of the present report shows that the requirement to self-register any publications based on ESSB data can be made much clearer to ESS users and proposes several ways in which to engage them more with the ESSB and to make registration easier. A significant number of additional publications are identified through the work of Dr. Malnar at University of Ljubljana. The work flow is analysed and recommendation for optimising this is presented. It is recommended to continue using Google Scholar as the main data source for this – despite its shortcomings, Google Scholar gives the best chance of broad coverage of publication types beyond journal articles.

A large number of smaller recommendations for improving the usability of the ESSB search pages are given. This includes e.g. adding faceted navigation in the search results, and to display the additional coded data from University of Ljubljana.

The potentials of bibliometric analysis in relation to ESSB is analysed. It is recommended to continue and extend the support for the work done at University of Ljubljana by Dr. Malnar to identify and analyse publications that use ESS data – the work is of very high quality and will be hard and expensive to automate given to relatively limited annual volume. For reliable estimations of relative citation impact on journal articles it is recommended to contract such analysis and interpretation from CWTS. Potentials in obtaining and displaying raw citation numbers of single publications based on Scopus and Google Scholar is also analysed. It is likely that updated citation numbers can be retrieved and displayed automatically without subscription costs.

Finally, the collaboration between NSD and University of Ljubljana is analysed. The comprehensive coverage of the ESSB is the result of the symbiotic work relation between the two sites. A number of recommendations are given to strengthen this collaboration and to enhance it, including directly comparing and linking the publications identified at both sites in order to identify duplicates, discover missing publications and make it possible to make the additional coded data from University of Ljubljana searchable and displayable in the ESSB.
6 Acknowledgments
The report rests in no small part on the knowledge provided in interviews and emails by Knut Kalgraff Skjåk and Ole-Petter Øvrebø from the Department for Survey and Data Services, NSD and Brina Malnar from University of Ljubljana.

7 References


## Appendix A – ESSB data registration steps

### ESS Bibliography

**Step 1: Authors, Title etc.**

- Select type of publication:
  - Journal article
  - Book (monograph)
  - Book chapter (part of anthology)
  - Report, working paper
  - Thesis, dissertation
  - Conference paper/poster
  - Newspaper/magazine article
  - Available manuscript
  - Anthology/Edited Volume

- Add authors:
  - Surname: 
  - First name(s): 
  - Institution: 
  - Country: Country not specified

- Title and language of publication:
  - English title:
  - Original title:

- Abstract (in English)

### Using this form:

*Fields marked with a red asterisk (*) are required.*

**Authors:** Start writing author's surname. As you write, the system may suggest names already registered in the database. Then enter the name and country of the institution with which the author was mainly affiliated when writing the publication.

To add co-authors, click "more" and repeat process.

**Abstract:** The publication should be summarised in English, preferably in 200-500 words. Avoid abbreviations, diagrams and references to the text.

**Type of publication:** Select one of the publication types. On the next page you will be asked to fill in further information specific to that type. A description of each type will appear when you move the mouse cursor over it.

---

*Figure 14. Step 1 - Information about publication type, author(s), title, language and abstract.*
**Journal**

<table>
<thead>
<tr>
<th>Name:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Print ISSN:</td>
<td></td>
</tr>
<tr>
<td>Online ISSN:</td>
<td></td>
</tr>
</tbody>
</table>

**Other details**

- *Refereed:*  
  - No
  - Yes

- *Year:*       
- *Volume:*    
- *Issue:*     
- *From page no:*  
- *To page no:*  
- *DOI Name:*    

---

**Figure 15.** Step 2 - Information about the source, dependent on publication type. For instance for journals: journal name, ISSNs, peer review status, year, volume, issue, page numbers and DOI.
Figure 16. Step 3 - Information about data use, that is ESS round/year, and which countries were used. Also contact email and any other comments.
### Figure 17. Step 4 - Review and confirmation of entered data – with possibility to edit the provided information.

<table>
<thead>
<tr>
<th>Author(s):</th>
<th>test test (Did you mean...), test, Country not specified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original title:</td>
<td>test</td>
</tr>
<tr>
<td>English title:</td>
<td>test</td>
</tr>
<tr>
<td>Language:</td>
<td>German</td>
</tr>
<tr>
<td>Type:</td>
<td>Journal article</td>
</tr>
<tr>
<td>Abstract:</td>
<td>test</td>
</tr>
</tbody>
</table>

**ESS data from (if applicable):** Round 1: Albania

**Contact e-mail:** test@test.com

<table>
<thead>
<tr>
<th>Year:</th>
<th>1234</th>
</tr>
</thead>
<tbody>
<tr>
<td>From page no:</td>
<td>1234</td>
</tr>
<tr>
<td>To page no:</td>
<td>1234</td>
</tr>
<tr>
<td>Volume:</td>
<td>1234</td>
</tr>
<tr>
<td>Issue:</td>
<td>1234</td>
</tr>
<tr>
<td>Refereed:</td>
<td>No</td>
</tr>
<tr>
<td>Journal:</td>
<td>test (Did you mean...)</td>
</tr>
</tbody>
</table>

Please type the registration code shown in the picture above:

[Test Registration Code](Try a different image)

[Confirm]
Appendix B – Possibilities in the Google Scholar, Publish or Perish and duplicate checkers

9.1 Avoiding the top1000 result limit in Google Scholar by Boolean searches

As noted in Section 4.1.2, a new challenge reported by Dr. Malnar is that the number of hits in recent years for a given year of ESS publications now exceed 1000 hits before the removal of duplicates. Due to Google limitations the consequence is that only the first 1000 documents presented by Google can be retrieved. As the ranking criteria of Google Scholar are only partially known this introduces uncertainty in which publications are included in the download and which are excluded.

Although Google is not a true inverted file system or relational database it does support Boolean operators. These can be utilised to attempt avoiding the top1000 problem by splitting the result set into smaller and hopefully non-overlapping sets. The overall idea is that when there are more than 1000 results, one or more auxiliary search terms are added to the query, and then excluded from the query. The aim is to split the large result set into smaller ones.

For example, for Google Scholar search that has less than 2000 results (here limited to the year 2012):  

Search #1: "European Social Survey"  \(\rightarrow\) About 1,390 results

Search #2: "European Social Survey" Germany  \(\rightarrow\) About 928 results  \(\rightarrow\) download

Search #3: "European Social Survey" -Germany  \(\rightarrow\) About 507 results  \(\rightarrow\) download

The total number of results exceeds the initial search (1435 results) and it is likely there is a small overlap between the two new sets. This method allows to create two sets that (most likely) contain all of the initial results – and that can be downloaded. Search #2 requires that both the phrase “European Social Survey” and “Germany” be present somewhere in the documents. Search #3 excludes all documents with the term “Germany”.

For a Google Scholar search of more than 2000 and less than 3000 results, an additional step is needed (here limited to the year 2016):

Search #4: "European Social Survey"  \(\rightarrow\) About 2250 results

Search #5: "European Social Survey" -Germany  \(\rightarrow\) About 848 results  \(\rightarrow\) download

Search #6: "European Social Survey" Germany Denmark  \(\rightarrow\) About 782 results  \(\rightarrow\) download

Search #7: "European Social Survey" Germany -Denmark  \(\rightarrow\) About 760 results  \(\rightarrow\) download

Search #5 retrieves all ESS results excluding the ones with the term “Germany”. Search #6 retrieves the ones with both “Germany” as well as “Denmark”. Search #7 retrieves the remaining ones – the ESS results with “Germany” but excluding the ones with Denmark. See Figure 18 for an illustration. Each set is in this case less than 1000 and can be downloaded. Again, the total (2390) exceeds the initial Search #4 and it is likely that there are small overlaps between Search #5, #6 and #7.

---

All searches executed on August 9, 2018 and with Google Scholar set as follows: Search for articles, excluding patents and search only for articles in English, and with citations included.
The selection of the auxiliary search terms is a process of trial and error. The terms have to fairly frequent in the documents, but in-frequent enough to not be in all of them. The aim is have each result set to be less than 1000 results so they can be downloaded.

Figure 18. Venn-diagram illustrating how to split a set of between 2000 and 3000 Google Scholar results.

9.2 Using Publish or Perish (PoP)

An alternative to the web interface of Google Scholar is the Publish or Perish (PoP) program. Created by Professor Anne-Wil Harzing PoP it facilitates the retrieval and analysis of citation data from Google Scholar and other sources. PoP has the same limitation in that it can only retrieve 1000 records in a single search, but all of these can be exported as a single file – reducing the need for much manual clicking in the initial download phase.

PoP is a Windows-based open source software that can be used free of charge once installed. The main interface for Google Scholar searches can be seen in Figure 19. It includes a query history, query fields, results window with publications and summary statistics.

Once a query is entered and ‘Lookup’ is clicked the query starts running. Depending on the query, on the current load on Google Scholar and on previous recent use this may take up to several minutes to complete. By a quick comparison to searches executed in the Google Scholar web interface, it seems that all languages are searched, and that citations are included (this is publications where Google Scholar has not indexed the full-text, but only found the citation in other full-texts) – these settings cannot be changed in PoP. Once a query completes the retrieved documents can be downloaded in a structured format in one go.

Similar tactics as presented in Section 9.1 can be used to limit the number of results to less than 1000 (see Figure 19). Not that it is not necessary to wait for the query to complete to know if a search will results in

---

23 See Harzing (2007) and https://harzing.com/resources/publish-or-perish
more than 1000 results – the expected number of results is given in the query window (that is 899 in Figure 19).

Figure 19. Publish or Perish software main window with query in progress.

9.3 Using reference manager duplicate checkers

Most reference manager software contain checkers that can identify potential duplicates in their databases\(^{24}\). These will search their respective databases for possible duplicates and offer to merge or delete them. Duplicates are found by matching on several document fields, e.g. authors names, title, source (journal/conference/publisher) etc. The matches are typically non-exact, that is, data with variations in spelling and even missing words can be tolerated and leads to a match. The exact implementation depends on the specific software product.

In relation to the ESSB potential duplicates could be identified by 1) exporting data in a format that the chosen software can import, and 2) running supplicate detection, and 3) using this to merge or delete duplicates in the ESSB. This could be done for instance:

- Quarterly on the entire ESSB publication list to identify any potential duplicates already in the ESSB.
- On data downloaded from Google Scholar or Publish or Perish to identify and remove duplicates at an early stage.
- At the time of import of data from the University of Ljubljana to the ESSB and NSD – as a first pass to identify records already in the SSB (i.e. self-reported publications) and separate them form new

\(^{24}\) See for instance:
- [https://www.refworks.com/rwsingle/help/Finding_Duplicate_References_In_Your_Account.htm](https://www.refworks.com/rwsingle/help/Finding_Duplicate_References_In_Your_Account.htm)
- [http://libhelp.ncl.ac.uk/faq/173447](http://libhelp.ncl.ac.uk/faq/173447) (for EndNote)
- [https://www.zotero.org/support/duplicate_detection](https://www.zotero.org/support/duplicate_detection)
publications. This can also be used to attempt to match semi-automatically ESSB records to all
Google Scholar publications and coded data from University of Ljubljana.

Note that the reference manager duplicate checkers should be relied on the provide 100% correct matches
in all cases – the results need to be verified.
Appendix C – Bibliographic control

Figure 20. Several potential duplicate records in the ESSB and lack of bibliographic control (e.g. Breen, M vs. Breen, Michael) – highlight added.
**Figure 21. Lack of bibliographic control on publisher (Oxford Univ. vs. Oxford University).**
Figure 22. Two different entries for the same publisher (Academia) – clicking on the publisher name in the full records does not lead to the same publisher entry.