CrossMark

Advancing research data publishing practices for the social sciences: from archive activity to empowering researchers

Veerle Van den Eynden[1] · **Louise Corti**[1]

Abstract Sharing and publishing social science research data have a long history in the UK, through long-standing agreements with government agencies for sharing survey data and the data policy, infrastructure, and data services supported by the Economic and Social Research Council. The UK Data Service and its predecessors developed data management, documentation, and publishing procedures and protocols that stand today as robust templates for data publishing. As the ESRC research data policy requires grant holders to submit their research data to the UK Data Service after a grant ends, setting standards and promoting them has been essential in raising the quality of the resulting research data being published. In the past, received data were all processed, documented, and published for reuse in-house. Recent investments have focused on guiding and training researchers in good data management practices and skills for creating shareable data, as well as a self-publishing repository system, ReShare. ReShare also receives data sets described in published data papers and achieves scientific quality assurance through peer review of submitted data sets before publication. Social science data are reused for research, to inform policy, in teaching and for methods learning. Over a 10 years period, responsive developments in system workflows, access control options, persistent identifiers, templates, and checks, together with targeted guidance for researchers, have helped raise the standard of self-publishing social science data. Lessons learned and developments in shifting publishing social science data from an archivist responsibility to a researcher process are showcased, as inspiration for institutions setting up a data repository.

✉ Veerle Van den Eynden
  veerle@essex.ac.uk

1   UK Data Archive, University of Essex, Colchester, UK

1 Introduction

The sharing and publishing of research data in the social sciences has a long history in the UK. Already in 1967, a data archive was established by the then Social Science Research Council at the University of Essex in the UK, to be able to safeguard valuable survey data from getting lost and make them available for secondary use [1]. From the 1970s, the governmental statistical service enabled government surveys, such as the General Household Survey, the Labour Force Survey, and the Family Expenditure Survey, to be deposited with this archive. The research council also invested strongly from the 1960s in the creation of large surveys, such as the British Election Survey, the British Household Panel Survey, Understanding Society, etc. Data resulting from such key surveys were equally disseminated via this archive. Where in the early days, data were disseminated as punch cards and magnetic tapes, with a printed catalogue publishing the available datasets, this evolved to a computerised web-based catalogue from 1994 onwards, and online data download from 2000 onwards. Still to this date, large and longitudinal surveys created by government departments, and research institutions are the most in demand amongst the data sets disseminated by the UK Data Archive, although the range of data now available is very diverse, comprising qualitative and quantitative data resulting from a diversity of research methods [2]. The top 20 downloads for longitudinal surveys represent half of all user downloads (37,231 of a total of 76,675 downloads over the last year) (Fig. 1). The three most in demand survey data series, the
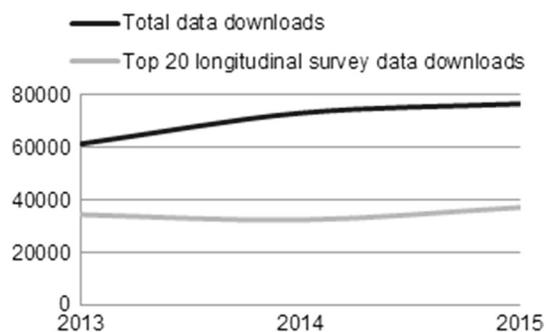
⚛ Springer

Fig. 1 Total annual data downloads and combined downloads of top 20 longitudinal surveys

Quarterly Labour Force Survey, the Health Survey for England, and the British Social Attitudes Survey, account for one fifth of all downloads. In contrast, a review of all downloads of qualitative and mixed methods data for the period 1994–2013 showed 5000 user downloads for 566 data sets [3].

From the 1990s, the investment by the Economic and Social Research Council (ESRC)—the main public funder for social sciences research in the UK—into data infrastructure was complemented by investments into holistic data services: the Economic and Social Data Service (2003–2012), followed by the UK Data Service (2012–2017), managed by the UK Data Archive at the University of Essex and Jisc Manchester. These support services provide guidance, advice, and training to data creators, data depositors, and data users.

From the mid-1990s, the ESRC adopted a research data sharing policy mandating data sharing as a condition of research funding. Research grant holders are expected to deposit the data that result from their research project with the UK Data Service, to enable their future reuse for research and learning. Since 2011 ESRC also requires a data management plan to be submitted with grant applications. All seven UK research councils have jointly adopted Common Principles on Data Policy [4]. The core principle is that publicly funded research data should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property. The most recent edition of the ESRC research data policy [5] aligns with these common data sharing principles, with practical guidelines for their implementation. In addition, the policy outlines roles and responsibilities of all actors in the research data landscape.

Overall, the activities of the UK Data Service are guided by the UK Strategy for Data Resources for Social and Economic Research [6], which ensures that decision making and policies based on social science-based evidence is as robust as any other science.

This paper showcases the developments made over time by the UK Data Service, and lesson learnt in shifting the publishing of social science research data as being an archivist-

controlled activity towards empowering researchers to do this themselves, based on developed data standards and flexible infrastructure provisions, with quality controls carried out by the data service or by peer reviewers.

2 Social sciences data reuse and value

Data resources of the UK Data Archive are being reused for research and to inform policy, but also to analyse and develop research methods and for teaching. For example, Poortinga et al. [7] compared public perceptions of climate change and energy futures between Britain and Japan before and after the Fukushima accident, because of existing data from public perceptions surveys carried out in Britain and Japan before [8] and after [9] the accident. Vogl et al. [10] show how data from the Health Survey for England [11] could be analysed to inform public health strategies, in particular providing evidence that smoking negatively affects health-related quality of life. Regarding research methods, Ogden and Cornwell [12] show by analysing 400 interview questions and their corresponding responses from 10 qualitative studies in the area of health that the richness of interview data can be predicted by open questions, being located later on in an interview and being framed in the present or past tense. Quah [13] shares his experience of using the IMF Direction of Trade Statistics, IMF Balance of Payment Statistics, and World Bank World Development Indicators in teaching students about understanding the global economy using real-world data. In addition, Bishop [3] found that by analysing 5000 downloads of qualitative and mixed methods data sets from the UK Data Archive between 1994 and 2013, nearly two-thirds of downloads are by students and over 60 % of uses are indicated as being for teaching. She also found that nearly all (96 %) of 566 collections have been used at least once. Usage metrics are currently based on downloads of data, since data citation and the use of digital object identifiers (DOIs) are not yet established enough to enable tracking data reuse via publications.

A recent independent review of the value and impact of well-established research data centres in the UK, including the UK Data Archive, showed that they have a large measurable impact on research efficiency and on return on investment in the data and services [14]. For the UK Data Service, the benefit/cost ratio of net economic value to operational costs was 5.4–1, and the increase in returns on investment in data and related infrastructure arising from additional use facilitated by the service was up to 10–1.

3 Social sciences data standards

Throughout its existence, the UK Data Archive has worked with and refined robust processes and standards for the cura-

Table 1 Data processing procedures at the UK Data Archive, comparing in-house data processing procedures with current status of self-publishing by researchers

In-house processing	Researcher self-publishing—archive staff task	Researcher self-publishing—researcher task
Review data for disclosive information	Yes	
Check data files for basic inconsistencies in data		Yes
Carry out data enhancement to agreed standards	Provide instructions for researchers	Yes
Generate the collection in multiple file formats for dissemination and preservation	Provide recommended file formats guidance; check formats are suitable	Yes, upload files in recommended formats
Generate variable and code list or datalist	Provide instructions and examples	Yes
Collate and prepare user documentation as bookmarked PDF/A documents	Review completeness of documentation	Include documentation files in random format in data collection package
Create enhanced DDI-compliant catalogue metadata for Discover catalogue, from information provided by the depositor in the deposit form and in documentation files	Review completeness and sufficient detail	Repository system metadata entry form facilitates submission of DDI-compliant catalogue metadata
Gather citations to related publications for inclusion in the catalogue metadata		Include links to publications in catalogue metadata
Assign a Digital Object Identifier (DataCite DOI) to the data collection	System automatically assigns DataCite DOI	
Release data via the UK Data Service Discover system	Publish after review and checks	

tion and publishing of the data it receives from researchers and data producers (Table 1). The international community of social science data archives has had shared approaches to preparing digital data for dissemination for over 40 years, building on a common descriptive vocabulary, the Standard Study Description Scheme, that was agreed back in 1976 by the Council of European Social Science Data Archives (CESSDA). A shared and common methodology that meets longer term curation needs ensures that data are independently understandable and remain preserved and accessible in the long term.

At the UK Data Archive, all data received for in-house processing are first assessed for disclosure risk, to ensure that studied individuals or organisations cannot be identified from the data, where this has been promised at the consent stage. Then, data integrity, missing values, and any anomalies or inconsistencies in the data are checked, as well as the file formats used, to ensure that they are the optimal format for long-term preservation and dissemination. Finally, the quality and composition of descriptions and documentation are examined, to ensure that the context of data provided is meaningful to a new user. The types of enhancements to data that may be carried out during processing are described in Table 2.

Files are processed according to the standard procedures [15,16]. Variables and code lists are generated for survey data sets and a datalist (an item-level finding aid) for qualitative data collections. Documentation files supplied by the depositor that include documents, such as questionnaire forms, interview question lists, sampling and fieldwork reports, and other descriptions of methods and context are grouped into a single user guide or a series of topical documentation guides that will complement the data set [17]. A structured metadata record is created that captures key descriptive attributes of the study and resulting data, using the Data Documentation Initiative (DDI) metadata standard; an example of which is the catalogue record for the British Social Attitudes Survey [18].

The DDI is a rich and detailed metadata standard for social, behavioural and economic sciences data, used by most social science data archives in the world [19,20]. DDI records contain mandatory and optional metadata elements relating to study description, data file description, and variable description. The study description elements contain information about the context of the data collection, scope of the study (e.g., subject topics, geography, time, method of data collection, sampling, and processing), data access information, information on accompanying materials, and provides a cita-

Table 2 Data enhancement procedures at the UK Data Archive, comparing in-house data processing procedures with current status of self-publishing by researchers

In-house data enhancements	Researcher self-publishing—archive staff task	Researcher self-publishing—researcher task
Add or edit variable and value labels in survey data files	Review for completeness	Yes
Convert interview transcripts to the standard UK Data Service template, including speaker tags and header of identifying information	Provide instructions, template and examples	Yes
Improve and harmonise file names according to data collection event	Provide instructions	Yes
Convert data files to preservation format	Review format	Yes
Convert data files to dissemination format	N/A	N/A
Prepare and publish key survey data and metadata to online Nesstar system	N/A	N/A
Prepare and publish high profile qualitative data and metadata to online QualiBank	N/A	N/A

tion. The data file description indicates data format, file type, file structure, missing data, weighting variables, and software used. Variable-level descriptions indicate the variable labels and codes. Initially, the DDI standard developed along the lines of a traditional social science codebook (DDI Codebook), while the recent version (DDI Lifecycle) focuses on the life cycle and reuse of data and metadata [21].

For indexing the study, the Humanities and Social Sciences Electronic Thesaurus (HASSET) is used as a controlled vocabulary to assign keywords that are tagged to the data sets [22]. The resulting package of data and documentation files, after conversion to suitable formats, is then placed both on the preservation systems, as well as onto online dissemination systems. Some high profile data are further published to online data browsing systems, such as Nesstar [23] for large-scale surveys, and QualiBank [24] for significant qualitative collections. Both systems allow users to dig down into the data sets exploring and analysing the data online. Via Nesstar users can explore, cross tabulate, visualise, and analyse individual variables and questions of survey data sets (Nesstar). QualiBank searches the content of text files, such as interviews, essays, open-ended questions and reports, as well as related metadata, such as descriptions of images and audio recordings, and enables hyperlinking to related objects. It also allows users to cite an entire data set or just extracts. Additional intensive processing work is carried out to enhance such studies. For survey data destined for Nesstar, the text of all questions asked during the survey and questionnaire routing information is added as metadata for the respective variables. For qualitative data destined for QualiBank, text is checked for typos and, where available, is linked to related sources, such as audio recordings for interviews and photographs. Similar data processing and cataloguing procedures are followed at other national social sciences data archives around the world.

The advantages of depositing data with a specialist data repository may include: assurances that data meet set quality standards; long-term preservation of data in the standard and accessible file formats; safe-keeping of data in a secure environment with the ability to control access where needed; online resource discovery of and access to data through data catalogues; front-line user support; and promotional and training opportunities for the data collection offering greater visibility in the data landscape.

Increasingly social sciences data sets are also published in data journals, such as Scientific Data, Research Data Journal for the Humanities and Social Sciences, or the Journal of open psychology data, whereby a data paper describes the data generation methodology, provenance, and reuse potential for a data set lodged in a repository in detail.

4 Publishing research data: from archive activity to DIY

In the early days of publishing academic data, all data collections received by the UK Data Archive were processed, documented, and prepared for reuse in-house. This activity can be prohibitively expensive. An analysis of the long-term costs of digital preservation for research data across eleven UK and two European data archives showed that the costs of acquisition, ingest, and access activities far outweigh the cost of archival storage and preservation. For the UK Data Archive, the cost of ingest (preparing and processing data sets for ingest into the archive) represents about 20 % of the total archive cost and is the most expensive step in the archiving process [25]. This meant that the number of data sets that could be curated, archived, and published on a yearly basis was limited, with a selection made of data collections on offer (Fig. 2). With increasing research funding

Fig. 2 Number of ESRC grants ending and resulting published data sets (not all ESRC grants generate research data)
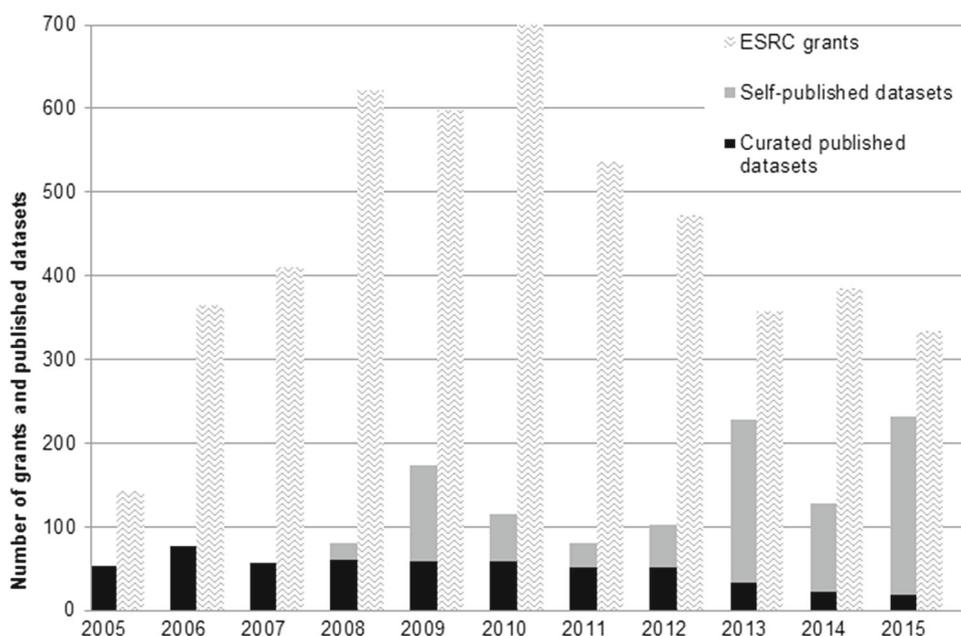
in the social sciences, there was a desire by the ESRC to see all data resulting from research grants equally and fairly archived and available for reuse. Technical advances also make it easier for researchers themselves to undertake data publishing activities. In addition, the original data creator (the researcher) has a better understanding of the research data, so, while it is still time-consuming to properly format and prepare data and add metadata, the data creator can accomplish these tasks in less time than would be required by a data archive curator who does not know the data in depth. Consequently, about a decade ago the archive started investing more in proactively guiding, training, and supporting researchers in good data management practices and skills for creating shareable data, as well as developing a self-publishing data repository system with prescriptive guidance and instructions so researchers can curate and publish data to the established archival standards. The repository system uses a DDI-compliant metadata profile aligned with the archive profile

The result of this concerted activity under the banner of research data management services is a collection of the best practice guides, handbooks, and accompanying teaching materials on relevant research data management topics (Table 3), following the logic of the data lifecycle [26–28]. This is complemented by extensive online guidance on the UK Data Service website and a programme of regular training workshops ranging from short introductory webinars or 2-h face-to-face sessions, to advanced 2-day hands-on courses, for diverse audiences of doctoral students, senior researchers, research support staff, and research managers. The guidance includes various examples and exercises developed from real data collections, as well as templates

Table 3 Topics of research data management guidance for researchers in the social sciences

Importance of sharing data

Consent, confidentiality and ethics of data sharing

Rights associated with the use of existing data

Describing, contextualising, and documenting research data

Data formats and software

Storage, back-up and security in response to the fragility of digital data

Publishing data in a sustainable way and enabling future citation

researchers can use, such as a template consent form that takes data sharing into consideration, a transcription template for transcribing interviews, and a datalist template for collections of qualitative data items.

In early 2014, the newly developed ReShare self-deposit data repository [29], an extensively customised version of Eprints open-source repository software, replaced its predecessor the fedora-based ESRC Data Store, and became the primary publishing system for social sciences research data in the UK, including data resulting from ESRC grants (Fig. 2). ReShare enables researchers to easily self-publish collections of research data and to make them available for use by other researchers. Its features include an easier-to-use depositor interface and more intuitive workflow (Fig. 3) than its predecessor. Design was influenced by the Eprints workflow commonly used by many libraries for their output repository. ReShare further simplifies the deposit of data sets by enabling the upload of multiple files in zip bundles, multiple data types, and associated documentation files. The ease of use is evidenced by the repository manager who cor-

responds one-to-one with most depositors experiencing far fewer queries about problems or confusion over the upload system. Data publishing usually proceeds without intervention from the repository manager, apart from the quality checks carried out.

The repository metadata profile is based on the DDI schema and aligns with the UK Data Service profile, whereby the workflow makes is easy to submit the necessary metadata elements in a step-by-step process. Customised-controlled vocabularies are aligned with those used in the UK Data Service's Discover portal. Access control options allow researchers to make data available to users as open or safeguarded data, and a DOI is attached to each deposit, so researchers can cite and track their own data collections. The data collections are discoverable via the Discover portal of the UK Data Service, amongst its portfolio of 7000 data collections.

The repository provides practical and easy guidance (Fig. 3) for researchers on preparing and documenting data files before deposit and publication, based on the exten-

sive in-house expertise that results from years of assessing, processing, and documenting social sciences data collections. It also shows the data review procedures (Fig. 3) that UK Data Service staff will carry out once data are submitted and before they are published [30], as visible indication of our expectations.

By July 2015, ReShare contained about 800 published data collections, spanning qualitative and quantitative research data. Reviewing this vast volume of self-published data has enabled us to identify the common problems researchers may face when publishing their research data (Table 4), adapt guidance, and provide solutions to avoid such common problems in the future. On the whole, the ReShare deposit experience is found to be positive for most depositors, with mostly good quality data and documentation being uploaded and shared.

In general, we handle such problems by relaying submitted collections back to the depositor for editing; by reiterating the quality expectations for data, metadata, and documentation files; by improving help guidance and directing depositors to

Fig. 3 Snapshot of the ReShare depositor workflow when describing and uploading a data collection, with circles indicating guidance and data review procedures

Table 4 Common problems encountered with self-publication of research data, and how to remedy them

Problem	Why problematic	Solution
Limited or poor key descriptive metadata in the catalogue record such as abstract or methodology description	Users have difficulty in understanding the exact content of the data when looking for data to use. Scientific reuse of data is difficult when key methodology information is absent	Return the data record to the depositor to add richer and more detailed information. Emphasise importance of detailed metadata via video tutorial and webinars
Poor or lack of contextual documentation	Users have difficulty understanding and interpreting the data	List of mandatory documentation to be uploaded for different types of data is given, e.g. Readme file, survey questionnaire, interview questions, data list.
Poor file naming, using special characters and not describing content	File names are incomprehensible	Include file naming guidance and request edits of file names where needed.
Data files not in recommended preservation formats	Risk of data files becoming redundant and inaccessible in the long term	Accept only data files in recommended preservation formats. Return the dataset to the depositor to convert to recommended formats

it; and by improving in-system checks, such as input controls. We have also started showcasing excellent collections on the ReShare home page as exemplars for future depositors and to give credit for best practice.

The overall result of the guidance and training for researchers and the self-publishing infrastructure development, with continued development of guidance and system in response to issues raised by data depositors, is that we are achieving many of the in-house data processing and data enhancement procedures to be carried by self-publishing researchers, whereby instructions are provided and checks done by archive staff (Tables 1, 2).

We provide succinct guidance on how to prepare data collections for self-publishing and which measures to take to produce well-documented collections suitable for long-term curation, both in the help guidance, and within the system workflow. Practical suggestions are flagged up when starting the deposit process, as well as at the stage of uploading data and documentation files (Table 5).

Therefore, by providing an easy-to-use, step-by-step self-publishing system, complemented by detailed data management guidance online and in best practice guides, together with a regular programme of training workshops for researchers, we can empower researchers to develop their data management skills. We can then focus our own expertise on quality assurance of the published data by reviewing each data set before publication. This involves checking for good levels of metadata and documentation, and ensuring they conform to ethical and legal requirements. In addition, we liaise with researchers prior to data deposit, to allay their concerns, and to answer the questions they have. This is often related to ethical concerns over data publishing.

Table 5 Advice for preparing a data collection for deposit given at the start and during the data deposit process

Group data files in zip bundles (max 2gb) according to their content or file format
For large collections, keep a folder structure for files in zip bundle
Check our recommended file formats before uploading files
Give files meaningful names that reflect the file content, avoiding spaces and special characters
Check that data files contain no disclosive information, with five suggestions on how to anonymise files
Create a ReadMe file for your data collection, with four suggestions for content
Prepare essential documentation to upload with data, with a list of essential items of documentation to include

In line with recent developments in the data publishing world, ReShare also receives data sets described in published data papers, such as Scientific Data, and facilitates peer review of submitted data sets prior to their publication for scientific quality assurance [31]. This means that at the review stage between a depositor submitting a data set and the publishing of this data set, peer reviewers selected by the journal are given access to the data set to review the data set itself for research quality. This complements the checks we carry out ourselves for the quality of documentation and metadata, and disclosive information in data. Only after reviews have been completed, any required edits to the data set done by the depositor and the journal publishes the data paper, is the data set published. Enabling such an innovative peer review of data required system changes to provide peer reviewers access to unpublished data records, the agreement of procedures with journals, and staff guidance on the handling of the peer review process.

5 Skills

Enabling researchers to be able to deposit high-quality data ready for publication and reuse into a data repository requires them to gain or enhance data management and data handling skills, in the topical areas listed earlier (Table 3). This can be gained through the kind of webinar and face-to-face training we provide, or via online learning modules, such as the Mantra research data management training [32]. Ideally, such data management skills training becomes part of the standard undergraduate or postgraduate research methods training [33].

For those individuals tasked with managing and administering a data repository, proactive engagement with researchers, research institutions, and research funders to achieve this goal requires technical knowledge and research skills. The presence of and familiarity with skilled support services helps data creators be less wary about data sharing mandates, and encourages positive collaboration with various research centres. Activities of the data repository manager may involve opening and understanding the content of files, handling data and quality assessment, disclosure review, and appreciating the data collection methods used and assessing technical documentation.

At the UK Data Service, most staff that engages with research data publishing has postgraduate research training, and those who provide training have extensive research expertise. Without hands-on research expertise, one would struggle to appreciate the challenges of sometimes complex fieldwork situations, technical research protocols, and ethical concerns over data publishing, and how to bring data sharing and data publishing into established research practices. Ideally, staff can embrace qualitative and quantitative research methods. Technical skills needed relate to metadata and data standards. A successful data publishing setup also needs good leadership with international connections, an eye for innovation, and the ability to collaborate when funding gets tough.

A data repository equally needs data users that are skilled in applying analytical methods, and in particular, the potential and pitfalls of secondary data analysis. In the UK, research methods training does focus on such skills, and we can see more courses embracing data management skills, and thus unifying the skills required for both data creation and data reuse.

6 Conclusion

The research data sharing and publishing landscape evolves rapidly worldwide, driven by technical advances, as well as research needs and expectations of funders, publishers, and governments with regard openness, transparency, and efficient investment of research. In the social sciences, the long-standing expertise of the UK Data Archive in curating, preserving, and publishing valuable data sets is increasingly applied to enhance data publishing practices and the associated skills of researchers. The well-established data management, data documentation, and data publishing procedures are applied to advice and train researchers in this area, so data sharing opportunities can increase. This is augmented by infrastructure and support service developments that empower researchers to self-publish their data to the standards and expectations of the current research and data publishing environment. These combined efforts raise the standard of self- publishing social science data and serve as example to institutions developing their own data repository. Guided by feedback, comments, and queries of researchers self-publishing their social science data, and innovations such as peer review for data journals, the UK Data Archive has fine-tuned its system workflow, instructions, and review procedures, to advance data publishing, and with it the availability of rich data resources for research and learning. As a next step, we will formalise quality ratings for submitted and published data sets, based on our review criteria [30], to give further credit to high-quality data sets, and inspire new depositors to meet these standards.

References

1. UK Data Archive (2007) Across the decades: 40 years of data archiving. http://data-archive.ac.uk/media/54761/ukda-40thanniversary.pdf. Accessed 26 June 2015
2. Corti, L.: Recent developments in archiving social research. Int. J. Soc. Res. Methodol. **15**(4), 281–290 (2012)
3. Bishop, L.: Re-using qualitative data: a little evidence, ongoing issues and modest reflections. Studia Socjologiczne **3**(214), 167–176 (2014). http://www.data-archive.ac.uk/media/492811/bishop_reusingqualdata_stsoc_2014.pdf. Accessed 15 Nov 2015

4. Research Councils UK (2011) RCUK common principles on data policy. http://www.rcuk.ac.uk/research/datapolicy/. Accessed 30 June 2015

5. ESRC (2015) ESRC Research data policy. http://www.esrc.ac.uk/_images/research-data-policy_tcm8-34123.pdf. Accessed 30 June 2015

6. UK Data Forum (2013) UK strategy for data resources for social and economic research 2013–2018. http://www.esrc.ac.uk/images/UK%20Data%20forum%20strategy_tcm8-28338.pdf. Accessed 30 June 2015

7. Poortinga, W., Aoyagi, M., Pidgeon, N.F.: Public perceptions of climate change and energy futures before and after the Fukushima accident: a comparison between Britain and Japan. Energy Policy **62**, 1204–1211 (2013). doi:10.1016/j.enpol.2013.08.015

8. Pidgeon, N.F., Hulme, M., Poortinga, W., O'Riordan, T., Lorenzoni, I.: Public risk perceptions, climate change and the reframing of UK energy policy in Britain, 2005. Data collection, UK Data Service. (2006). doi:10.5255/UKDA-SN-5357-1

9. Pidgeon, N., Spence, A., Poortinga, W.: Public perceptions of climate change and energy futures in Britain, 2010. Data collection, UK Data Service (2010). doi:10.5255/UKDA-SN-6581-1

10. Vogl, M., Wenig, C.M., Leidl, R., Pokhrel, S.: Smoking and health-related quality of life in English general population: implications for economic evaluation. BMC Publ. Health **12**(203) (2012). doi:10.1186/1471-2458-12-203

11. National Centre for Social Research, Department of Epidemiology and Public Health University College London (2011) Health Survey for England. Data Collection, 4th edn. UK Data Service (2006). doi:10.5255/UKDA-SN-5809-1

12. Ogden, J., Cornwell, D.: The role of topic, interviewee and question in predicting rich interview data in the field of health research. Sociol. Health Illn. **32**(7), 1059–1071 (2010). doi:10.1111/j.1467-9566.2010.01272.x

13. Quah D.: Understanding the global economy using real-world data. Case study, UK Data Service (2011). https://www.ukdataservice.ac.uk/use-data/data-in-use/case-study/?id=18. Accessed 15 Nov 2015

14. Beagrie, N., Houghton J.: The value and impact of data sharing and curation: a synthesis of three recent studies of UK research data centres. JISC Report (2014). http://repository.jisc.ac.uk/5568/1/iDF308_-_Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf. Accessed 15 Nov 2015

15. UK Data Archive (2014) Quantitative data ingest processing procedures. http://www.data-archive.ac.uk/media/54770/ukda081-ds-quantitativedataprocessingprocedures.pdf. Accessed 25 June 2015

16. UK Data Archive (2014) Qualitative data collection ingest processing procedures. http://www.data-archive.ac.uk/media/54767/ukda093-ds-qualitativeprocessingprocedures.pdf. Accessed 26 June 2015

17. UK Data Archive (2014) Documentation ingest processing procedures. http://www.data-archive.ac.uk/media/54785/ukda078-ds-documentationprocessingprocedures.pdf. Accessed 26 June 2015

18. NatCen Social Research: British social attitudes survey. Data Collection, UK Data Service (2013). doi:10.5255/UKDA-SN-7500-1

19. Blank, G., Rasmussen, K.B.: The data documentation initiative: the value and significance of a worldwide standard. Soc. Sci. Comput. Rev. **22**(3), 307–318 (2004). doi:10.1177/0894439304263144

20. Vardigan, M., Heus, P., Thomas, W.: Data documentation initiative: toward a standard for the social sciences. Int. J. Digit. Curat. **3**(1), 107–113 (2008)

21. Vardigan M.: The DDI matures: 1997 to the present. IASSIST Q. **37**(1–4), 45–50 (2013). http://www.iassistdata.org/sites/default/files/iq/iqvol371_4_vardigan.pdf. Accessed 26 June 2015

22. UK Data Archive: HASSET thesaurus (2015). http://www.data-archive.ac.uk/find/hasset-thesaurus. Accessed 26 June 2015

23. UK Data Service (2015) UK Data Service Nesstar browser. http://nesstar.ukdataservice.ac.uk/webview/. Accessed 26 June 2015

24. UK Data Service. UK Data Service QualiBank (2015). http://discover.ukdataservice.ac.uk/QualiBank. Accessed 26 June 2015

25. Beagrie, N., Lavoie, B., Woollard, M.: Keeping research data safe (Phase 2) (2010). http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf. Accessed 15 Nov 2015

26. Van den Eynden, V., Corti, L., Bishop, L., Woollard M.: Managing and sharing data: best practice for researchers. UK Data Archive, Colchester (2011). http://www.data-archive.ac.uk/media/2894/managingsharing.pdf. Accessed 30 June 2015

27. Corti, L., Van den Eynden, V., Bishop, L., Morgan B.: Managing and sharing data: training resources. UK Data Archive, Colchester (2011). http://repository.essex.ac.uk/2398/1/TrainingResourcesPack.pdf. Accessed 30 June 2015

28. Corti, L., Van den Eynden, V., Bishop, L., Woollard, M.: Managing and Sharing Research Data: A Guide to Good Practice. Sage, London (2014)

29. UK Data Service. ReShare (2014). http://reshare.ukdataservice.ac.uk. Accessed 30 June 2015

30. UK Data Service. ReShare data review procedures (2015). http://reshare.ukdataservice.ac.uk/reshare-review-procedures. Accessed 30 June 2015

31. UK Data Service. Scientific Data approves UK Data Service as recommended data depository. Impact Blog (2015). http://blog.ukdataservice.ac.uk/scientific-data-approves-uk-data-service-as-recommended-data-depository. Accessed 30 June 2015

32. EDINA, Data Library. Research data MANTRA. Online course, University of Edinburgh (2011). http://datalib.edina.ac.uk/mantra. Accessed 30 June 2015

33. Corti, L., Van den Eynden, V.: Learning to manage and share data: jump-starting the research methods curriculum. Int. J. Soc. Res. Methodol. (2015). doi:10.1080/13645579.2015.1062627