Responding To Needs, Driving Change

CDI
CHILDHOOD DEVELOPMENT INITIATIVE

Sharing Social Research Data in Ireland
A Practical Toolkit

Dr Bríd McGrath and Robin Hanan • 2016

How to cite this report

Any citation of this report should use the following reference:
McGrath, B. and Hanan, R., Sharing Social Research Data in Ireland: A Practical Toolkit (2016) Dublin: Childhood Development Initiative (CDI).

Published by Childhood Development Initiative, Dublin

ISBN: 978-0-9928763-7-1

Disclaimer

While every care is taken to ensure that this report is as up-to-date and accurate as possible, no responsibility can be taken by CDI for any errors or omissions contained herein. Furthermore, responsibility for any loss, damage or distress resulting from adherence to any advice, suggestions or recommendations made available through this report, howsoever caused, is equally disclaimed by CDI.

Copyright

Acknowledgements

The authors are grateful to the following for their comments on earlier drafts of this Toolkit:

Julia Barrett, Research Services Manager, U.C.D. Library; Ruth Geraghty and Jane Gray, Irish Qualitative Data Archive, Maynooth University; Amanda Quail, Growing up in Ireland; Mark Ward, Trinity College Dublin and Suzanne Guerin and Marian Quinn, CDI.

Our deepest appreciation also to the evaluation teams whose experience informed this Toolkit. A very special thanks to Liam O'Hare at Queens University for showing us all it can be done!

CDI would like to acknowledge the support of our funders, the Department of Children and Youth Affairs, and The Atlantic Philanthropies. As always, they recognise the importance of legacy and longevity of learning, principles which underpin the archiving process.

Contents

Glossary of Terms

Data Collection	A data collection is the data lodged in an archive, and typically includes three or four components: dataset(s), documentation, Metadata and possibly code;
Data Integrity	Maintaining and assuring the accuracy and consistency of data, which is a critical aspect to the design, implementation and usage of any system which stores, processes, or retrieves data;
Digital Preservation	"*A set of activities required to make sure digital objects can be located, rendered, used and understood in the future.*" (www.digitalpreservationeurope.eu);
Harvesting Data	A technique for extracting metadata by automatic means from individual repositories and gathering it in a central catalogue to facilitate search interoperability;
Imputed Data	The substitution of estimated values for missing or inconsistent data items (fields);
Irish Research Council	The body set up in 2012 as a merger of the Irish Research Council for Humanities and Social Sciences (IRCHSS) and the Irish Research Council for Science, Engineering and Technology (IRCSET);
Metadata	Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource;
Non-proprietary software	Free or open-source software;
Primary Documentation	The original material created during the study, regardless of format, e.g. datasets, information sheets, consent forms, ethical approval;
Secondary Documentation	Documentation which is derived from or ancillary to the primary documentation or which supports or explains it;
Secondary use	The use of data collected by someone else for further research;
Tertiary Documentation	Articles, books and other similar materials which use the primary and secondary documentation.

Further terms are defined in http://researchdata.ox.ac.uk/home/glossary/

Acronyms

CDI	Childhood Development Initiative
CESSDA	Consortium of European Social Science Data Archives
DDI	Data Documentation Initiative
IQDA	Irish Qualitative Data Archive
ISSDA	Irish Social Science Data Archive
RCTs	Randomised Controlled Trials;
RECSAF	Research Ethics Committee Standard Application Form
SPSS	A Statistical Package for the Social Sciences, a software package used for statistical analysis. It is now officially named “IBM SPSS Statistics”

Introduction

Why should you share your data collections?

Your data is valuable and has an importance outside your own original project. Allowing other researchers to reuse your data maximises the impact of your work, and benefits both the scholarly community and society in general. Sharing your data allows other researchers to use your material in ways you may not have thought of, or may not have been able to do within your research project. It allows other researchers to replicate your findings, to verify your results, test your instruments and compare with other studies. It also allows them to use your work to expand knowledge in important areas. It provides value for money by reducing duplication and advancing knowledge and also has a significant value in education, as it allows both graduate and under-graduate students to develop their skills in qualitative and quantitative research by using high-quality data in their studies, without having to conduct their own surveys.

Archiving your data also guarantees its long-term preservation and accessibility. As many research teams are assembled only for individual projects, long-term preservation and access to research data collections can only be guaranteed if they are deposited in an archive which will manage them, ensure access and provide user-support. In addition, the archives will ensure that the datasets do not become obsolescent or corrupted.

Finally, increasingly funders require that you make your research data available as a condition of their funding your research, so that other researchers can test your findings, and use your data to extend research in your area. Equally, publishers are also specifying access to research data as a condition for publication.[1]

What data can you share?

Apart from your findings, you can share any of your own instruments (tests and scales), code books, information sheets, consent forms and ethical approval.

How can you archive your data?

There are many ways of archiving your data; you can archive on your own server, or deposit your data in a data archive. For most practical purposes, it is preferable to deposit the datasets in an open access data archive, which will preserve and manage the datasets, ensure access to *bona fide* researchers, provide support to users and track the secondary use of the datasets. This is especially relevant for small organisations which do not have the resources to manage the datasets or provide support, or where research teams are dispersed after the research project is completed.

This Toolkit has grown out of the experience of the authors in advising the Childhood Development Initiative (C.D.I.), Tallaght, on the archiving of materials from a number of research projects. It is written primarily with Irish researchers and those who commission research, in mind, but it is hoped that it will be useful in other contexts also.

We also draw on a developing literature on good practice. We have not referenced these in much detail here, but we would refer readers to our *Report on the archiving of CDI Data*.

---

[1] http://www.nature.com/authors/policies/availability.html Accessed 6 January 2016.

Archiving data requires specific skills which research teams may need to acquire. The resources listed below provide very useful guidance, and the staff of data archives are also available to support research teams in the planning and implementation of data archiving plans.

http://dri.ie/digital-archiving-in-ireland-2012.pdf,

http://dri.ie/digitisation-resources

http://www.nuigalway.ie/hbsc/documents/lp__key_issues_data_strategy_nov_2011.pdf

http://libguides.ucd.ie/data/

http://www.iqda.ie/content/planning-archive-your-postgraduate-research-data-6-things-bear-mind

and Corti, L., Eynden, V. V. D., Bishop, L., & Woollard, M. (2014). *Managing and sharing research data: a guide to good practice*.

Since the Toolkit is intended as a practical guide in an ever-evolving field, the authors will be grateful for useful feedback which can improve the usability and accuracy of future editions.

As the Toolkit has grown out of C.D.I.'s research projects, the various examples given here relate to children; however the same considerations apply to any cohort of research participants, especially those who may be vulnerable.

What to do if you decide to share data late in the project

Ideally, research teams should plan to share their material before beginning their projects; early planning makes later sharing much easier, and also less costly.[2]

In many cases, projects start without research teams being aware that data sharing would be a useful or necessary element in increasing the value of their data.

This can mean that consent forms, information sheets, ethical guidelines, and record formats and management were not designed with archiving, and the many practical, ethical and legal issues which it throws up, in mind. Many research teams also have limited practical experience of digital archiving and open-access datasets, and should certainly talk to the archives' data managers early in their projects.

Given that current thinking about archiving suggests that additional uses may be found for data, it may be regarded as more ethical to make the material available for secondary use, provided that participants' privacy is safeguarded by anonymising the data, and that ethical approval is granted for such secondary use. Where data is anonymised, there is no infringement of the Data Protection Legislation. The Irish Government's recent decision to retain historical data from heel-prick tests without seeking retrospective consent provides a useful precedent in such cases, but one which must be approached cautiously, as this Toolkit spells out.

This means that a researcher or research institute must weigh up the potential risk of archiving the quantitative and qualitative data, which is the possibility of some of the participants or their parents or guardians being offended or upset by the material being made available, even where their own individual information is not

---

[2] http://www.data-archive.ac.uk/create-manage/planning-for-sharing/costing accessed 6 December 2015.

traceable. If the researcher is satisfied that this is a slight risk, then it is possible to archive the quantitative datasets and some of the qualitative material, provided that ethical approval is obtained.

It has often been perceived that anonymising **qualitative** data can be difficult, but recent developments and new resources have made it much easier. However, as with all datasets, heavy anonymisation also runs the risk of reducing the value of the data. This can be compensated for by arranging different levels of access to all or part of a dataset, as is done, for example, by the Central Statistics Office.

Questions to consider in late decisions to share data

1. Who owns the data?
2. Is the data sufficiently important to warrant making it available without explicit consent? (see, for example, the decision not to destroy blood samples collected in the Irish *Heel Prick Tests*, as discussed in http://www.hse.ie/eng/health/child/newbornscreening/newbornbloodspotscreening/disposal/disposalqa.html Health Service Executive (2014, November).
3. Can you sufficiently anonymise the data to protect privacy and confidentiality?
4. Can you obtain retrospective consent to share data?
5. Is this best done by having an opt-in option?
6. Can you obtain retrospective ethical approval?

Developing a data management plan

If you wish to make your datasets available to future researchers, whether in your own organisation or in a public archive, you should ideally develop a data management plan before the start of a research project and integrate it into the aims, objectives, actions and budget of the project. In particular, ensure that your consent forms, information sheets and ethical approval include permission to archive data and allow for possible secondary use, and that you have funding for the work involved in depositing the data. Where you may need to impute data, ensure that you have a master file of the original data stored and available for later archiving.

The governing body of the research organisation should take responsibility for the archiving plan. Responsibility for implementing the plan should be delegated to a named individual, and all researchers and research teams need to be aware of their specific responsibilities.

Prior informed consent of participants (or, where appropriate the parents and guardians of minors) is a pre-requisite to archiving the datasets.

The key decision is whether to retain the materials in your own organisation or to deposit them in a public archive.

In the latter case, you will need to identify appropriate repositories for different types of material and to discuss their ethical standards, formats for materials to be transferred and practical arrangements.

Formal agreements with the repository (ies) should ensure appropriate preservation of data, access by researchers and the public, and conditions of use of the data. Ideally, this should be agreed before the start of the research project and discussed, and revised as necessary, as the project continues.

Develop a detailed plan for maintaining and sorting data for future archiving. Budget for the costs of storing, sorting, preparing and transferring data. You may need two separate plans, a basic one for your funders, and a longer, living document to ensure efficient and effective working practices for the research team.[3]

Some of the issues to be addressed in the plan are listed in Table A and discussed in more detail on the following pages.

Table A

Elements of a Data Management Plan

1 • What materials and datasets will be generated by the project?
2 • What can and should be made public?
3 • Do you have the informed consent of participants?
4 • Do you have the ethical approval to make data available?
5 • Do you want to make materials available in your own organisation or in a public archive?
6 • Which archive is appropriate for each type of material?
7 • What are the requirements of the archives(s)?
8 • What formats will you need?
9 • How will you ensure data integrity?
10 • How will you manage data protection?
11 • How will you anonymise the data?
12 • How do you need to format and prepare materials for transfer?
13 • What metadata will you need?
14 • How will you promote use of the archive?

---

[3] http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements Accessed 6 December 2015.

1. What materials and data will be generated by the project?

As part of your planning for the research project, you should map out what data will be generated by the project. This will need to be updated throughout the project.

It can be useful initially to divide the materials into categories such as:

- Quantitative materials, both primary and secondary;
- Code books, instruments;
- Primary qualitative materials;
- Secondary (summary) qualitative materials;
- Administrative or management data, including consent forms, ethical approval;
- Tertiary material, such as reports, articles, conference papers etc.

You should list the various categories of information in terms of file types and formats, as well as predicted quantities of data, as indicated in Appendix I.

• Quantitative Data (both Primary and Secondary)

This can include:

- Datasets and the material from which they were developed, such as questionnaires, score sheets, etc., and
- Associated ancillary and administrative material – e.g. manuals, code-books, interviewer instructions, consent forms, information sheets, ethical approval, Garda (police) vetting, etc.

Quantitative data files can more easily be made available to the public as long as they are clear, consistent, anonymised raw data. This will often be in summary (secondary) form.

If you need to manipulate the data, or impute some data, ensure you keep a master file of the original data, for reference and archiving.

• Primary Qualitative Data

The primary qualitative data can include soundfiles, transcripts and reports of focus group and individual interviews, field notes, survey responses, etc. Qualitative data presents different challenges as anonymising the data can be time consuming. For a useful guide, see https://content.web.nuim.ie/sites/default/files/assets/document/IQDA_Best_Practice_Handbook_0.pdf Accessed 6 December 2015.

• Administrative and Management Data

These include progress reports, minutes, newsletters, blog posts and other administrative documentation. In many cases, these may not be suitable for archiving or necessary to retain.

• Tertiary Material

This will include studies derived from the data, apart from the main reports, and could include journal articles, conference papers, presentations and subsequent publications. This should be listed with links to the secondary publications where available.

2. What can and should be made publicly available?

Researchers need to think carefully about the types of materials to be produced by a particular project and whether they are suitable to be made publicly available.

We have already indicated that the trend is now towards making such materials available where possible. However, this can only be done if it does not compromise the rights and privacy of the research participants.

In the case of each category or group of files, researchers should consider:

- What is the potential value of the data gathered during the research to future researchers and/or policy makers? How can this be enhanced?
- Is it possible to anonymise the data sufficiently to protect the privacy of participants while retaining its utility?
- The balance between the value of making this material available to future researchers and the costs and complexity of anonymisation.

3. Do you have ethical approval to share the materials?

Data sharing is dependent on appropriate ethical approval. This should ideally be obtained at the start of the research project (See Appendix II). However, as already mentioned, many research projects may not have considered this issue at the start, so it may be necessary to obtain ethical approval retrospectively before making datasets public.

Ethical approval decisions should be guided by legal requirements, good practice and the research organisation's own policies on subjects such as child protection, as well as the requirements of their Research Ethics Committee.

4. Do you have the informed consent of participants?

Prior informed consent of participants (or, where appropriate the parents and guardians of minors or others who may not be able to give their consent) is a pre-requisite to sharing the datasets. This is essential to protecting their privacy and rights and also facilitates the smooth transfer of data to the archives at the end of the project.

When planning the project, you should include explicit consent to archive the data in:

- The consent forms;
- The information sheets for participants;
- The ethical approval for the project.

All participants, parents or guardians must be made fully aware of the meaning of the consent forms before agreeing to them. Therefore, consent forms and information sheets must be written in clear and straightforward language. You should also explain clearly how the data will be anonymised and who will be able to access it.

Requesting consent to archive data may influence people’s willingness to participate in research, or at least affect the nature of that participation. However, it is unavoidable if the material is worth archiving.

Sample consent forms and a covering letter are given in Appendix IV, V and VI, but these will need to be adapted to the particular study and target group.

Retrospective Consent

Where explicit consent has not been obtained in advance, it may still be possible to archive the data by getting retrospective consent from participants. In some cases where the datasets have been thoroughly anonymised, it may be possible to seek retrospective ethical approval (see Appendix III), provided that nothing in the consent forms and information sheet specifically exclude this.

5. Do you want to share the materials in your own organisation or in a public archive?

It is possible to archive your research collection within your own organisation, or through a hosted service, such as Amazon’s https://aws.amazon.com/public-data-sets/.

However, your data will receive more attention, and is more likely to be found and reused if it is held in either of the main Irish data archives, http://issda.ucd.ie/ or https://www.maynoothuniversity.ie/iqda. In addition, the archives’ staff will maintain the data collections, provide and manage access and deal with all the administrative work involved in hosting your data collections.

6. Which archive is appropriate for each type of material?

If you choose not to retain the material yourself, or dispose of it, it is important to choose an archive, or an archive for each type of data, which is/are appropriate to the different elements of your study.

The archiving, preservation and control of the data collections requires a team:

- With a thorough understanding of good practice and ethical issues in digital archiving;
- With the appropriate technical and administrative skills;
- Which can be relied on to maintain and update the storage media, manage access and provide user support.

The institutions should also work towards accreditation from an approved body, such as the Consortium of European Social Science Data Archives (CESSDA), as these standards are adopted.[4]

CESSDA provides a listing of information resources on such archives in many European countries.

---

[4] https://www.ukdataservice.ac.uk/news-and-events/newsitem/?id=4419 accessed 18 January 2016.

In Ireland, the most important archives for social science data are:

- The Irish Social Science Data Archive of Quantitative Datasets in UCD;
- The Irish Qualitative Data Archive, NUIM, and
- The Northern Ireland Qualitative Archives (NIQA), currently being developed in QUB.

The UK Data Archive covers both qualitative and quantitative data. Some examples of datasets held by a range of other archives can be accessed here:

- Data Sets: biodiversityireland.ie;
- Data & statistics: National Cancer Registry Ireland;
- Ireland Datasets: Ryan Institute;
- Fingal Open Data;
- All-Island Research Observatory An Tionad Breathnaithe um Thaighde Uile-Éireann airo.ie.

7. What are the requirement(s) of the archive(s)?

Archives will have their own rules and requirements on what can be archived and under what conditions.

It is important to discuss, as early as possible, whether your data are suitable for the archive in question and what their particular requirements are. This needs to be reviewed continuously and built into a formal agreement with the archive(s).

This will allow the researchers to be clear on what needs to be done in relation to areas such as:

- Informed consent required from study participants to archive their information;
- Ethical approval to archive the data;
- Formats for storing data;
- Terms and conditions for access to the data;
- Anonymisation of materials **(see Stage 8)**.

This will also allow for clarity on the archives' policies on areas such as:

- Preservation of materials;
- Rules and supports for access by researchers and other interested parties.

The agreement between the archives and the researchers should include development of:

- **An end user licensing agreement** outlining how and for what purpose the data can be used (See Appendix VII). This must take into account the risk associated with each type of document. This should include a provision that users lodge a copy of the published results of their secondary research in their own organisations' repositories or another open repository, where they do not have their own repository; inform the data archive repository and supply a copy to the archives themselves. It should also specify how the datasets are to be cited for accountability, for the convenience of future users and to promote awareness and use of the archives. The **UK Data Archive End User Archive**

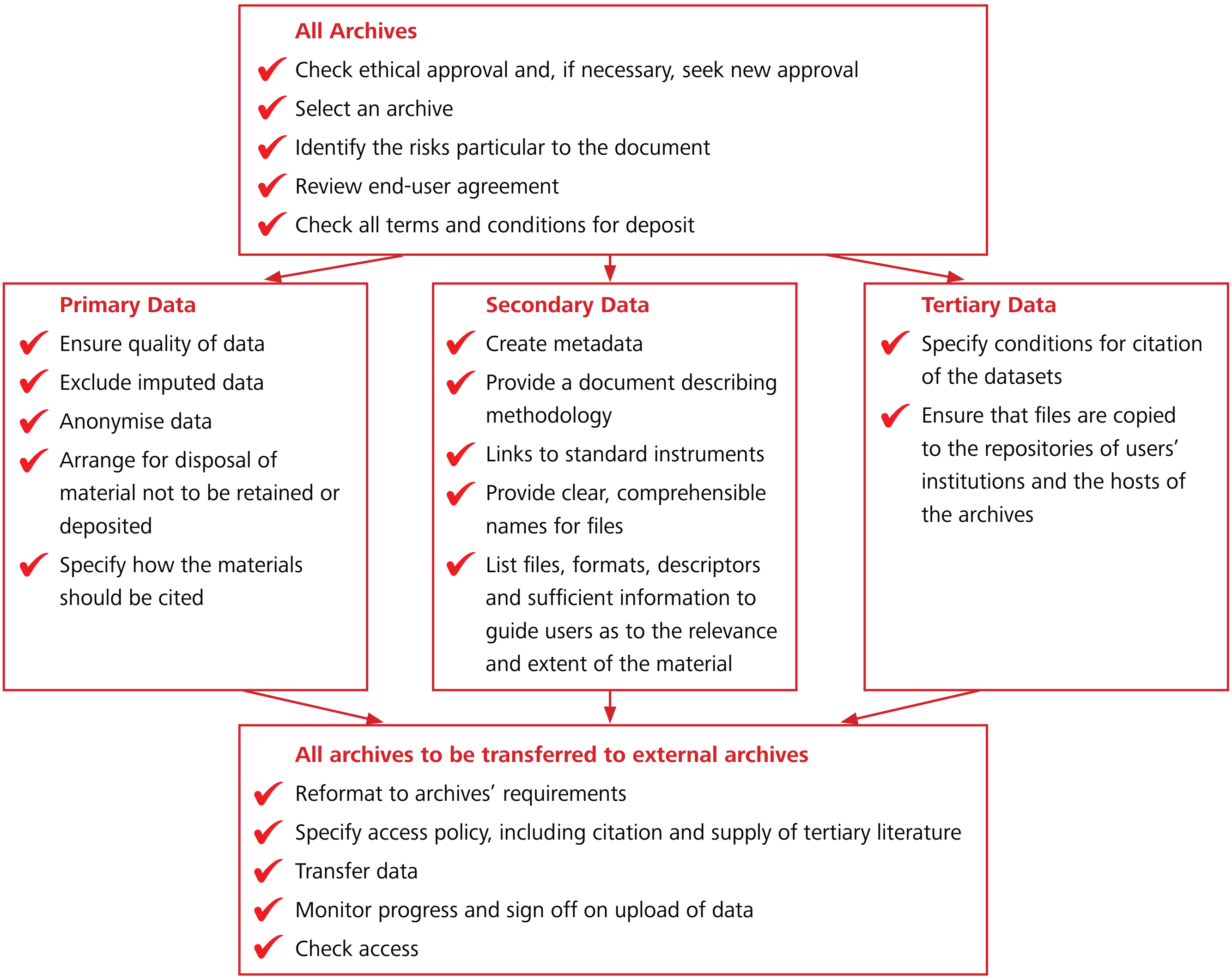
Agreement provides a useful example. Here is an example of how to cite a dataset.

- **An access agreement** covering levels of access and access restrictions. In the case of quantitative data, allowing everyone access is appropriate but for some data it may be more appropriate to allow full access to some data, and restricted access to some very detailed data, or for commercial use, e.g. the C.S.O.'s data sets. The archive must put in place appropriate systems to manage different access levels.
- Here is an example of general access to C.D.I.'s dataset from Doodle Den.
- An application form for researchers to apply to use the data archives. See, for example, the Irish Qualitative Data Archive data request form.
- **A Dissemination Information Package (DIP)** specifying relevant standards and processes for deposit, storage, preservation and dissemination (see Section II of this Toolkit).

It is also essential to reach agreement on who owns the intellectual property (copyright) of the data, any research tools or instruments and the study's results. The copyright owner can control what use is made of the material and how it may be disseminated. Copyright may be owned by the sponsoring organisation, the researchers or the funder, or a combination of any of these. If you do not own the copyright, you should get prior agreement on who may use the research materials and how.

Table B

Checklist on finalising data files for sharing



All Archives

✓ Check ethical approval and, if necessary, seek new approval
✓ Select an archive
✓ Identify the risks particular to the document
✓ Review end-user agreement
✓ Check all terms and conditions for deposit

Primary Data

✓ Ensure quality of data
✓ Exclude imputed data
✓ Anonymise data
✓ Arrange for disposal of material not to be retained or deposited
✓ Specify how the materials should be cited

Secondary Data

✓ Create metadata
✓ Provide a document describing methodology
✓ Links to standard instruments
✓ Provide clear, comprehensible names for files
✓ List files, formats, descriptors and sufficient information to guide users as to the relevance and extent of the material

Tertiary Data

✓ Specify conditions for citation of the datasets
✓ Ensure that files are copied to the repositories of users' institutions and the hosts of the archives

All archives to be transferred to external archives

✓ Reformat to archives' requirements
✓ Specify access policy, including citation and supply of tertiary literature
✓ Transfer data
✓ Monitor progress and sign off on upload of data
✓ Check access

8. How will you anonymise the data?

Sensitive data must be anonymised to the greatest extent possible. This is essential to ensure that individuals or groups cannot be identified, as required by good practice and by data protection legislation. Effective anonymisation can be time-consuming, but ensuring careful file management can reduce the work involved.

The anonymisation process involves both direct and indirect means of identification and is different for qualitative and quantitative data.

Direct identifiers include names and addresses, and these can easily be anonymised; **indirect identifiers** include health conditions, occupations and, for quantitative data, the cell size and gender. In those cases, anonymising the data to avoid inadvertent identification is essential.

For **quantitative** data, the general rule of thumb is that cell sizes of 20 or more ensures that data is sufficiently anonymised. For example, where only nine schools are involved in a data collection, the gender of a school principal can identify that person, or a rare health condition can identify a child. In these cases, great care must be taken to anonymise the data but in some data collections, this may not be entirely possible.

Sound files, which can provide a rich source of information to future researchers, are particularly difficult to anonymise.

Good practice in anonymising data of this type involves:

- Deletion of **direct identifiers** data (e.g. names, address, organisation);
- Removal of **indirect identifiers** which allow a person to be easily identified (occupation, health conditions, etc.);
- Replacing these with descriptions congruent with the subject matter.

A ‘tracking table’ can be kept to record changes and to link real names with pseudonyms, but this becomes problematic when the material is archived. Linking real names with pseudonyms or codes can lead to problems.

A paper on *Qualitative Research In Ireland: Archiving Strategies and Development* by Dr. Jane Gray and Dr Aileen O’Carroll summarises best practice and gives practical advice on this. In Ireland, anonymisation will follow the guidelines from the Irish Social Science Data Archive and the Irish Qualitative Data Archive and the UK Data Archive’s guidelines on anonymisation are also useful in this regard.

There is always a danger of over-anonymising data, and researchers sometimes want access to the original forms, to be able to take account of variables hidden by the anonymising, such as geographical data. ‘Data transformation files’ should be retained, but with the same security and access protocols as for the most sensitive original data. In many cases it will not be possible to provide access to this level of personal data without infringing privacy and data protection standards and laws.

In other cases, it may be appropriate to provide either for different levels of access to data-sets, for example to established researchers rather than graduate students, or to permit access only on-site and under supervision.

9. How do you format and prepare materials for transfer?

Data archives naturally expect the data they receive to be technically and legally suitable for use, adequately documented and relatively error-free. You should discuss the formats in which files can be archived with the relevant archive(s), which generally have very specific requirements.

Host institutions generally insist that all raw data be "cleaned up". This would mean, among other issues, excluding imputed data and appropriately anonymising all data.

Retrospectively removing imputed data can be an enormous task, depending on the quantity of data available and needs to be budgeted and planned for; this can be avoided by maintaining a clean 'master file' and using only copies of this when it is necessary to impute data. The master file can be archived and the secondary files discarded after completion of the research project.

Digital objects require pro-active intervention to remain accessible. Software upgrades may not support legacy files; the industry may not produce compatible software; and software may be bought by a competitor and discontinued. Information contained in digital archives will cease to be accessible without digital preservation to manage otherwise obsolescent file formats.

Proprietary software quickly becomes obsolete and even inaccessible, so repositories specify acceptable formats, generally in non-proprietary file formats (for example PDF rather than Word for documents). Some repositories will insist that the same material should be provided in a number of different formats.

A paper for CESSDA discusses acceptable formats across Europe. Data formats are specified in general terms on the ISSDA and websites. The U.K. Data Archive also provides a comprehensive list of formats which are acceptable to it (reproduced from July 2014 at Appendix VIII).

10. What metadata will you need?

Depositing good quality research data is of limited value without ensuring that potential users are made aware of the contents of the data and its potential usefulness to research, and a system which makes the data easy to find.

Future researchers must be provided with sufficient descriptive data (metadata) to enable them to locate and estimate the potential value of (the researcher's) primary material for their research. For more information about metadata, see http://www.usgs.gov/datamanagement/describe/metadata.php and http://libguides.ucd.ie/data/doc_metadata (accessed 6 December 2015).

Metadata takes the form of explanatory guides which list, describe and explain the primary data and make it easy for users to identify the primary material to which they require access. In addition, apart from the already published evaluations, future studies, either currently being written by the research teams or which may emanate from access to the archive material, should also be captured and made publicly available as far as possible.

It is important to get the technical metadata right, so that the contents of each file are correct, retrievable and understandable to users. The metadata linked to the archives must be exceptionally clear, and the secondary data (descriptors, code books to explain variables etc., e.g., manuals, copies of questionnaires/scales, interviewers' instructions) must be clear, accessible and comprehensive.

Metadata must meet international standards, including being compliant with the D.D.I. standard and be easy to harvest. It is important to have regard to the OAI-PMH (Open Access Initiative for Metadata Harvesting), the protocol which guarantees and supports interoperability between digital archives, and will therefore enable researchers to locate these datasets.

This is a significant task as it requires creating appropriate and comprehensive secondary documentation for each archive file. The level of necessary detail will vary for each study.

Archives often provide guides or templates for metadata. Here is a guide from the Digital Repository Ireland on creating metadata http://dri.ie/sites/default/files/files/metadata-quality-control.pdf.

11. How will you promote use of the archives?

To be of broad benefit to Irish society, the material deposited should be used not only by professional researchers but also by policy-makers, opinion-formers, social and child-care practitioners, NGOs and the general public.

To promote this access, it is important that all researchers should be required to cite the dataset in their publications in an agreed format and that copies of their resulting publications should be deposited in the archives themselves as well as in other repositories required by law and good practice.

These published reports should be made available in their parent institutions' repositories. They should also be made available through subject-based repositories such as Lenus.ie and preferably also through such resources as Google Books and databases such as ERIC.

In addition, the datasets should be listed in Thompson Reuters Data Citation Index.

More traditional forms of publicity should be used including, but not limited to:

- E-mails to lists and journals of interested parties (researchers and research institutions, professional associations, government department, community organisations, academics, journalists etc.);
- Social media, such as Twitter, Instagram, Facebook and social media groups;
- Presentations and poster sessions at conferences, research seminars etc.;
- Providing teaching packs suitable for classroom use;
- Publishing blog posts, e.g. http://www.casatravis.org/blog/post/advocating-for-children-who-have-witnessed-family-violence (accessed 17 January 2016).

Increasingly, researchers are encouraged to publish articles describing their datasets, either in specialist journals, or in a journal specifically designed only to include such articles such as Data in brief from Elsevier.

Data archives will also do some publicity when new datasets, or waves of existing data collections are lodged.

12. How will you dispose of or retain data which cannot be transferred?

Research institutes may wish to retain some materials which are not suitable for sharing with the public. This might include copies of published materials, administrative and financial materials related to the research project and various internal and process documents.

All other materials, which are not suitable for sharing, should be destroyed. This will include original, non-anonymised survey forms and summaries and any other material which cannot be anonymised.

Table C

Checklists on disseminating the results of your research

Pre-research	• Clarify copyright, ensure consent/ethical approval to use material and archive datasets; • Plan dissemination; • Arrange funding, including dissemination costs; • Start research blog to talk about your research; • Ensure link from your website; • Compile mailing list (including email lists) for dissemination; • Join social media groups (e.g. interest groups on LinkedIn) to post about your research.
During the Research	• Use newsletters and blogs to provide progress reports on your research.
Post research	• Publish report and summary; • Launch report, include on website and newsletters; • Load it on website with simple stable URL; • Lodge copies in legal deposit and other libraries; • Self-archive in (Researchgate, Mendeley, http://www.ssrn.com/en/, https://www.academia.edu/,) as appropriate; • Lodge in repositories (including http://www.issuelab.org/home, http://www.lenus.ie/hse/); • Respond to call for papers for suitable conferences; • Load powerpoints on Slideshare and Prezi; • Find and publish suitable audiovisual clips (Youtube / Vimeo / Vine) and embed them in your website; • Use Journal Guide to identify suitable journals for your articles; submit articles to them; revise in line with readers' reports; • Again, archive these articles and provide links from your blog and website / tweet about them.

A useful toolkit for disseminating research is http://www.uregina.ca/arts/assets/docs/pdf/Dissemination-Toolkit.pdf

Appendix I: A template for assessing the potential to make materials publicly available

NB: This checklist should be monitored and developed during the lifetime of the project as plans become more concrete.

File name (or group of files)	Who are the study group?	Do they include vulnerable groups?	Informed consent of participants or their guardians?	Ethical approval, including for archiving?	Who is responsible for the data collection and storage?	Qualitative or Quantitative?	Format	Cell size	Where can they be archived? Own organisation or (name of archive)?	Consent to archive?	Notes

Appendix II: Standard Ethical Application Form

The following is an excerpt from the RECSAF Standard Application Form, which is used by a number of ethics committees in Ireland. This excerpt focuses on sections relevant to the archiving of data.

SECTION E: Data Protection

E1.1 (a) Will consent be sought for the processing of data?

Yes

E1.1 (b) If no, please elaborate.

Not applicable.

E2.1 Who will have access to the data which is collected?

The researchers will have access to the data in identifiable form (audio recordings of interviews). Only de-identified data will be made available to other researchers. The information sheet will note the plan to archive data and participants will be invited to consent to this element in the consent form.

E2.2 What media of data will be collected?

Audio recordings of interviews and hard copies of survey data.

E2.3 (a) Would you class the data collected in this study as anonymous, irrevocably anonymised, pseudonymised, coded or identifiable data?

Classifications of data will include:

Anonymous data that are returned with no identifiers in the first place;

Irrevocably anonymised data, which are returned with an identifier which is then removed and no identification key is retained;

Pseudonymised data that are returned with an identifier which is replaced with a made up name and would mostly apply to qualitative data;

Coded data that are returned with an identifier, which is replaced with an identification code but a ‘key’ is retained;

Identifiable data that are returned with an identifier and no attempt is made to anonymise that identifier.

E2.3 (b) If 'coded', please confirm who will retain the 'key' to re-identify the data?

If data is coded the identification key will be retained, and only made available to key study personnel.

E2.4 Where will data which is collected be stored?

In the first instance with the Principal Investigator, however ultimately de-identified data will also be archived in a recognised research archive.

E2.5 Please comment on security measures which have been put in place to ensure the security of collected data.

All computers are password protected and audio files will be encrypted.

E2.6 (a) Will data collected be at any stage leaving the site(s) of origin?

Yes

E2.6 (b) If yes, please elaborate.

Data will be collected in participants' homes and in the organisational setting. Data will be transported to the academic site for transcription. Finally de-identified data will be archived in a recognised research archive.

E2.7 Where will data analysis take place and who will perform data analysis (if known)?

Analysis for this project will take place on the academic site and will be performed by the principal investigator. As the data are to be archived, additional analyses will be completed in the future.

e2.8 (a) After data analysis has taken place, will data be destroyed or retained?

All data will be retained until the project is completed. At that point, audio recordings will be destroyed and only de-identified data will be retained.

E2.8 (b) Please elaborate.

We propose to retain the de-identified data for further analysis/teaching and submission to a recognised research archive. We feel this will maximise the learning from the research.

E2.8 (c) If destroyed, how, when and by whom will it be destroyed?

Audio recordings will be destroyed once the project is completed.

E2.8 (d) If retained, for how long, for what purpose, and where will it be retained?

We propose to retain the de-identified data indefinitely for further analysis and teaching. This will include submission to a recognised research archive. We feel this will maximise the learning from the research. The archives’ procedures will ensure that it will only be used for research/teaching purposes.

E2.9 Please comment on the confidentiality of collected data.

Data will be gathered in identifiable form but will be de-identified for analysis and archiving.

E2.10 (a) Will any of the interview data collected consist of audio recordings / video recordings?

Yes – Audio recordings

E2.10 (b) If yes, will participants be given the opportunity to review and amend transcripts of the tapes?

They will be offered the opportunity to review and remove information from transcripts.

E2.11 (a) Will any of the study data collected consist of photographs/ video recordings?

No

E2.11 (b) If yes, please elaborate.

Not applicable.

Appendix III: Letter seeking retrospective ethical approval for archiving data where this had not been sought at the start of the project

Date

Dear XXXXXXXXXX,

This letter marks an addition to our original submission for ethical approval granted (XXXXXX date and ref). Our Board has requested that we seek additional approval from our Ethics Committee in relation to data archiving as this was not explicitly dealt with in our earlier submission. We hereby request ethical approval for the passing on of anonymised quantitative data about (name of research project).

That information is as outlined and as approved in our previous submission (see attached documentation). The researchers intend to archive the anonymised data arising out of the study for the purposes of future research but all participants and settings would remain anonymous and unidentified. If such research did go ahead, it would involve the analysis of secondary data.

Therefore we,

a. Request ethical approval to pass anonymised quantitative data on (name of research project? Or type of data?) once the study ends.
b. Request ethical approval for (name of researcher) to archive that anonymised data for the purposes of future research.

I would appreciate it if this letter was considered in addition to the submissions previously made in relation to this evaluation and I look forward to hearing from you (either for more clarification or for the decision of the committee in relation to this matter).

Yours sincerely,

____

(Name) (Lead Researcher)

Appendix IV: Sample Consent Form for research participants

(Two questions on archiving highlighted in colour here.)

Researcher(s)' Name(s):	**Title:** (Name of team)
Faculty/School/Department:	
Title of Study:	
To be completed by the PARTICIPANT:	

Please circle the relevant answer

- Have you been fully informed/read the information sheets about this study? YES/NO
- Have you had an opportunity to ask questions and discuss this study? YES/NO
- Have you received satisfactory answers to all your questions? YES/NO
- Have you received enough information about this study? YES/NO
- Do you understand that you are free to withdraw from this study;
    - at any time,
    - without giving a reason for withdrawing,
    - without affecting your future relationship with the research organisation? YES/NO
    - Do you agree to take part in this study, the results of which are likely to be published? YES/NO
- Have you been informed that this consent form and all information you provide will be kept in the confidence of the researcher? YES/NO
- **Do you agree that the information from this study can be made available to future researchers, in a form which ensures your privacy and anonymity?** YES/NO
- **Have the plans to preserve the information from this study been explained to you clearly, and all of your questions answered?** YES/NO

Name of participant (Block letters): ______________________________

Signature of participant: ______________________________

Signature of Researcher: ______________ Date: ______________

Table V: Sample Consent Form for parents/guardians of minors who are participants

(Two questions on archiving highlighted in colour here.)

Researcher(s)' Name(s):	**Name of Research Team:**
Faculty/School/Department:	
Title of Study:	
To be completed by the PARENT/GUARDIAN:	

Please circle the relevant answer

- Have you been fully informed/read the information sheets about this study? YES/NO
- Have you had an opportunity to ask questions and discuss this study? YES/NO
- Have you received satisfactory answers to all your questions? YES/NO
- Have you received enough information about this study? YES/NO
- Do you understand that you are free to withdraw from this study;
    - at any time,
    - without giving a reason for withdrawing,
    - without affecting your future relationship with the researchers? YES/NO
- Do you agree to take part in this study, the results of which are likely to be published? YES/NO
- Do you agree to be contacted to take part in the focus group (see information sheet) YES/NO
- Have you been informed that this consent form and all information provided on you or your child will be kept confidential to the researchers? YES/NO
- **Have the plans to preserve the information from this study been explained to you clearly, and all of your questions answered?** YES/NO
- **Do you agree that the information from this study can be made available to future researchers, in a form which ensures your child's privacy and anonymity?** YES/NO

Child Name: ____

Child Date of Birth: ____

Signed by Parent: ______________________ Date: ______________________

Name in Block Letters: ______________________

Parent Address: ______________________

____

____

Parent Contact Number: ______________________

Signature of Researcher: ______________________ Date: ______________________

Appendix V: Sample Information Sheet for Parents of Children Participating in the Research – to be adapted to the particular circumstances

This is a valuable piece of research and the research team really hopes that you and your child will take part. The information gathered will be used to see how XXXXXXXXXXXXXXX and the information will be presented in general rather than focusing on any one child, family or childcare setting.

Children will be given gifts such as colouring pencils as a small thank you for taking part and all parents who agree to participate will be entered into a draw for a €250 supermarket shopping voucher as a token of our appreciation.

Children

- Children will be given an Identification number at the start of the research and they will be identified by this number rather than by their name from then on. The only people who will have access to child names are the Lead Researcher (name) and the fieldworker (name). Child data will not be stored with names so there will be no way for anyone apart from the Lead Researcher to tell which child data goes with which name.
- Information will be gathered on child development at three different points: at the beginning of preschool, once at the end of the first year and once at the end of the second year. If your child moves to primary school after one year, we would still like to keep them in the research and we would like to see them after one year of primary school.
- The assessments are designed to be child-friendly and children tend to enjoy them as they are meant to feel like playing a game and they only last about 20 minutes.
- Information on children will be gathered by fieldworkers in the preschool setting. The fieldworkers are highly trained to work with children and parents.
- (The research organisation) will give updates in their newsletter so that parents can be kept informed about the research.
- We regret that we cannot give parents their individual child’s results. Our fieldworkers are not trained to be clinical or educational psychologists therefore they are not qualified or permitted to report on individual children. If you are concerned about your child, a member of the preschool staff will have ideas about how to get some extra help for them.
- The Lead Researcher will also be a point of contact should parents have any additional questions or concerns about the research (see contact details below). Please feel free to call or email me any time.

Parents

- We will gather information through two face-to-face interviews with parents.This will happen at a time and place that suits parents; once at the beginning of their child’s preschool and once again after two years have passed. The interview lasts only 15 minutes and is about the general experience of parenting young children.

- All information given will be treated as confidential. It is possible that parents might provide information concerning a young person at risk while being interviewed. In this case the researcher has a duty of care to respond to any such information in line with the Children First National Guidelines for the Protection and Welfare of Children (2009).

Thank you for your valued participation

Signed: The Research Team:

Contact: XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Appendix VIII: File formats table from the UK Data Archive

Note: This table is derived from the UK Data Archive website, specifically http://data-archive.ac.uk/create-manage/format/formats-table where you can access more information.

Be aware that Irish data archive format requirements may be different and that guidelines may change from time to time.

Type of data	Acceptable formats for sharing, reuse and preservation	Other acceptable formats for data preservation
Quantitative tabular data with extensive metadata: A dataset with variable labels, code labels, and defined missing values, in addition to the matrix of data	SPSS portable format (.por); Delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) containing metadata information; Some structured text or mark-up file containing metadata information, e.g. DDI XML file.	Proprietary formats of statistical packages e.g. SPSS (.sav), Stata (.dta) MS Access (.mdb/.accdb).
Quantitative tabular data with minimal metadata: A matrix of data with or without column headings or variable names, but no other metadata or labelling	Comma-separated values (CSV) file (.csv) tab-limited file (.tab); Including delimited text of given character set with SQL data definition statements where appropriate.	Delimited text of given character set - only characters not present in the data should be used as delimiters (.txt); Widely-used formats, e.g. MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf) and OpenDocument Spreadsheet (.ods).
Geospatial data: Vector and raster data	ESRI Shapefile (essential - .shp, .shx, .dbf, optional - .prj, .sbx, .sbn); Geo-referenced TIFF (.tif, .tfw); CAD data (.dwg); Tabular GIS attribute data.	ESRI Geodatabase format (.mdb); MapInfo Interchange Format (.mif) for vector data; Keyhole Mark-up Language (KML) (.kml); Adobe Illustrator (.ai), CAD data (.dxf or .svg); Binary formats of GIS and CAD packages.
Qualitative data Textual	eXtensible Mark-up Language (XML) text according to an appropriate Document Type Definition (DTD) or schema (.xml); Rich Text Format (.rtf); plain text data, ASCII (.txt).	Hypertext Mark-up Language (HTML) (.html); Widely-used proprietary formats, e.g. MS Word (.doc/.docx); Some proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti.

Type of data	Acceptable formats for sharing, reuse and preservation	Other acceptable formats for data preservation
Digital image data	TIFF version 6 uncompressed (.tif)	JPEG (.jpeg, .jpg) but only if created in this format; TIFF (other versions) (.tif, .tiff); Adobe Portable Document Format (PDF/A, PDF) (.pdf); Standard applicable RAW image format (.raw); Photoshop files (.psd).
Digital audio data	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) but only if created in this format; Audio Interchange File Format (AIFF) (.aif); Waveform Audio Format (WAV) (.wav).
Digital video data	MPEG-4 (.mp4); Motion JPEG 2000 (.mj2).	
Documentation and scripts	Rich Text Format (.rtf); PDF/A or PDF (.pdf); HTML (.htm); Open Document Text (.odt).	Plain text (.txt); Some widely-used proprietary formats, e.g. MS Word (.doc/.docx) or MS Excel (.xls/.xlsx); XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0.

The Childhood Development Initiative
St. Mark's Youth and Family Centre
Cookstown Lane
Fettercairn
Tallaght, Dublin 24

Tel: (01) 494 0030
Fax: (01) 462 7329
E-mail: info@twcdi.ie
Web: www.twcdi.ie
Twitter: @twcdi

CDI
Responding To Needs, Driving Change

The Area Based Childhood Programme 2013–2017

An Roinn Leanaí agus Gnóthaí Óige
Department of Children and Youth Affairs

The Atlantic Philanthropies

CES Centre for Effective Services

pobal
government supporting communities