Report to
The Childhood Development Initiative
on archiving of C.D.I. data

CDI
CHILDHOOD DEVELOPMENT INITIATIVE

Meeting needs, making changes, improving outcomes

Meeting needs, making changes, improving outcomes

CDI
CHILDHOOD DEVELOPMENT INITIATIVE

Report to
The Childhood Development Initiative
on archiving of C.D.I. data

This report details the results of a study by Dr Bríd McGrath and Robin Hanan for
the Childhood Development Initiative (C.D.I.).

The Aim of the study is to *develop guidelines and procedures for C.D.I.'s archiving process which draw
on nationally internationally agreed best practice for both qualitative and quantitative data*

2013

CHILDHOOD DEVELOPMENT INITIATIVE

How to cite this report

Any citation of this report should use the following reference:

McGrath, B. and Hanan, R. (2013) Report to Childhood Development Initiative (CDI)
    on archiving of CDI data, Dublin: Childhood Development Initiative (CDI).

Contents

Acknowledgements

Definitions used in this report

"**Digital preservation** is a set of activities required to make sure digital objects can be located, rendered, used and understood in the future. This can include managing the object names and locations, updating the storage media, documenting the content and tracking hardware and software changes to make sure objects can still be opened and understood."[1] ; "**Digital preservation** combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time."[2] ;

Harvesting Data: a technique for extracting metadata by automatic means from individual repositories and gathering it in a central catalog to facilitate search interoperability;

Imputed Data the substitution of estimated values for missing or inconsistent data items (fields)[3];

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;[4]

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 some other purpose;

SPSS: Statistical package for the social sciences, a software package used for statistical analysis. It is now officially named "IBM SPSS Statistics";

Tertiary Documentation: Articles, books and other similar materials which use the primary and secondary documentation.

---

[1] http://www.digitalpreservationeurope.eu/what-is-digital-preservation/#96 (accessed 28th. March 2013.

[2] ALA (American Library Association) (2007). *Definitions of digital preservation*. Chicago: http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.pdf

[3] OECD Glossary of Statistical Terms - Data imputation Definition stats.oecd.org/glossary/ detail.asp?ID=3406

[4] There is a more detailed definition at NISO: *Dublin Core Metadata Element Set Approved*. Source: http://www.niso.org/news/ releases/PRDubCr.html (accessed on 14th June 2008). Quoted in Jiban K. Pal, Practical framework for harvesting standard metadata in digital repository. 7th International CALIBER-2009, Pondicherry University, Puducherry, February 25-27, 2009 © INFLIBNET Centre, Ahmedabadhttp://www.inflibnet.ac.in/caliber2009/CaliberPDF/14.pdf.

A. Background and Methodology

The aim of this study is:

"To develop guidelines and procedures for C.D.I.'s archiving process which draws on nationally agreed best practice guidelines for both qualitative and quantitative data."

The C.D.I.

The Childhood Development Initiative (C.D.I.) was set up in 2003 to support better outcomes for children in Tallaght West. Since then, C.D.I. has commissioned research reports from three academic institutions covering the following evaluations:

- *Early Childhood Care and Education and the Speech and Language Therapy Model*, Centre for Social and Education Research; Dublin Institute of Technology;

- *Doodle Den and Mate-Tricks*, Centre for Effective Education; Queens University Belfast (the C.E.E. is conducting a follow-up study of Doodle Den);

- *Healthy Schools Programme*, School of Nursing and Midwifery; Trinity College Dublin;

- *Community Safety Initiative, Restorative Practice Evaluation* and the overall *Process Evaluation*; Child & Family Research Centre, School of Political Science & Sociology, National University of Ireland, Galway.

These research projects have utilised a range of methodologies, including randomised controlled trials, quasi-experimental design and process evaluations. Datasets with key primary research material have been created in all of the above evaluations and used as the basis for the published reports. Evaluation and the dissemination of research findings have been integral to all of C.D.I.'s work from the start.

The Galway evaluations were not included in this study as the evaluation team does not intend to archive its data.

Changing approaches to archiving data

These evaluations did not start with an awareness that archiving would be a core objective, or the implications of this process. This meant that the consent forms, information sheets, ethical guidelines, and record formats were not designed with archiving, and the many practical, ethical and legal issues which it throws up, in mind.

It would also appear that there is limited practical experience of digital archiving and open-access datasets, although at least one evaluation team showed a broad appreciation of many of the questions involved. The question of data sharing was raised at a meeting of all the research teams in November 2009 and it was agreed to work towards making the anonymised quantitative data available for secondary use, as far as possible.

The research teams and C.D.I. were not unusual in this respect. At the time when these evaluations were planned and launched, few researchers, especially in the social sciences, considered the possibility of making their datasets and other material available to other researchers.

Nowadays, researchers in all disciplines are becoming increasingly aware of the value of making their historic datasets available, although this has been slower in the social sciences than in many other disciplines.[5] This has contributed to the development of increasing numbers of longitudinal studies and is reflected in national policy. For example, the Irish Research Council now requires all applicants for grant awards to show how their data will be made available to other researchers[6].

[5] U.K. Data Archive, Across the decades: 40 years of data archiving. (2007). http://data-archive.ac.uk/media/54761/ukda-40thanniversary.pdf accessed 15 August 2013.

[6] *"Whenever data is to be collected with the support of a grant awarded by the Council and/or partners, applicants must specify the means by which that data will be made available as a public good for use by other researchers, via the Irish Social Science Data Archive (ISSDA) or other appropriate channels."*, Irish Research Council Government Of Ireland, Research Project Grants Scheme, 2013, Terms And Conditions, http://www.research.ie/sites/default/files/irc_rpg2013_terms__conditions_final_0.pdf

This applies especially, but not exclusively, to randomised controlled trials. Making this material accessible and useable to other researchers can ensure maximum use of the datasets for further research, incidentally increasing the value for money of the original studies, and allowing for comparative studies. Data of this type is important to policy-making and further research.

There has been much discussion within the research community, especially amongst scientists, on the subject of making raw data (such as is found in datasets) widely available.[7] There has been a significant shift not merely towards making such data-sets widely available, but also to the development of national digital archives in which to store, manage and preserve this material. Data curators and researchers are now becoming more aware of the need to disseminate information about their holdings, to encourage greater use of the valuable data they hold. There is also an increasing recognition of the importance of process evaluation, so that important lessons about the conduct of research studies may be learnt and disseminated.

In general, the ability to archive depends on participants' consent, which should be obtained at the outset of any research project. Retrospective consent may also be sought, although this is more problematical because of, among other things, the difficulty in tracking participants and their parents and guardians and possible reluctance by parents and guardians to give consent retrospectively. Equally, ethical approval ought to be sought for future possible archiving, and making datasets and other material available for secondary use. Where primary

and secondary documentation may be archived and made widely available, sufficient promotion and publicity must be arranged to make other researchers aware of its existence and value.

Within the C.D.I., there has been an increasing recognition of the importance of archiving the datasets and other material and this was reflected in, for example, discussions at evaluation team meetings. To date, individual evaluation teams have not given sufficient attention to the means or implications of archiving the datasets. However, an important report on issues in archiving qualitative data was produced by the Irish Qualitative Data Archive, with the support of C.D.I., in 2010 and one of the teams has considered some of the issues to be addressed in archiving their materials.[8]

Archiving sensitive materials

The new emphasis on archiving social studies datasets and secondary and ancillary materials raises important ethical questions about the protection of the privacy and rights of participants in the research. Over the recent past, researchers in all disciplines have come to appreciate the importance of clear ethical standards in research, including the implications for potential archiving of data[9]. It is now more widely appreciated that informed consent should be sought from participants for future secondary use of any data collected in the course of a study, and of the need to protect participants' privacy. There is also a concern that the possibility that such material may be made more widely available could influence people's willingness to participate in research, or at least affect the nature of that participation.

7   e.g., Hrynaszkiewicz, Iain, and Douglas G. Altman. "Towards agreement on best practice for publishing raw clinical trial data." Trials 10.1 (2009): 17. http://www.biomedcentral.com/content/pdf/1745-6215-10-17.pdf accessed 7 April 2013, Uhlir, Paul F., and Peter Schröder. "Open data for global science." *Data Science Journal 6* (2007): OD36-OD53. http://prijipati.library.usyd.edu.au/bitstream/2123/2676/1/LegalFramework_Ch8.pdf, Cragin, Melissa H., et al. "Data sharing, small science and institutional repositories." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368.1926 (2010): 4023-4038. http://rsta.royalsocietypublishing.org/content/368/1926/4023.full Palmer, Carole L., Nicholas M. Weber, and Melissa H. Cragin. "Analytic potential of data: assessing reuse value." *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*. ACM, 2011. Tenopir, Carol, et al. "Sharing data: Practices, barriers, and incentives." *Proceedings of the American Society for Information Science and Technology* 48.1 (2011): 1-4.
8   Irish Qualitative Data Archive and Tallaght West Childhood Development Initiative, Best Practice in Archiving Qualitative Data. (2010). http://research.nuim.ie/sites/research.nuim.ie/files/images /IQDA%20Best%20Practice%20Handbook.pdf
9   *Ensuring the Data-Rich Future of the Social Sciences Gary King Science 331, 719 (2011); http://bidm.stat.fju.edu.tw:81/2013-研究所Data%20Mining/ DM2013-0305/science-2011-2-11-/Science-12.pdf accessed 16 August 2013.*

These questions are particularly sensitive for projects like C.D.I., since they were not considered either in overall planning of the Initiative or in the individual evaluations.

One of the teams did request an amendment to its ethical approval to allow for archiving. One of the team's consent forms specified that the information was to be gathered only "for the purposes of the research study" while the information sheets for another stated "Any information that we collect about your child will be held confidentially by the research team … and only used for research purposes." Moreover, after the issue of data sharing was raised in November 2009, all of the research teams were asked to change their information sheets for the second cohorts, to allow for the possibility of lodging the data in an archive. It is unclear to what extent this happened, but it was certainly not done by all of the teams. In at least one case, the data for the second cohort had already been gathered before November 2009. This raises the possibility that part, though not all, of the datasets could be archived.

While none of the information sheets, consent forms or ethical guidelines guaranteed that the data would be destroyed, or expressly precluded the possibility of archiving the data emanating from the studies, none of them explicitly or implicitly obtained permission to allow the material to be used by anyone other than the original research team. Moreover in some cases the evaluation teams had sought revised ethical approval for making their quantitative data available outside the research teams. This data may be archived, but perhaps not the qualitative data, because of its ethical approval and the difficulty of anonymising this material. It is important to note, however, that, under the contracts between C.D.I. and each of the evaluation teams, the data is owned by C.D.I.

These provisions may be insufficient authority to archive the material, especially for the qualitative data. Moreover, if, as in one case, consent was sought for this study only, then that implies precluding access for secondary studies, and retrospective consent should be sought from all participants for archiving.

It is accepted that this will require significant time and staff resources, without any guarantee that all the participants will agree to the archiving. Were C.D.I. to proceed to archive without this retrospective consent, the collectors of the data (i.e. the researchers) would be liable for any subsequent mis-use in any secondary study.

However, current thinking about archiving suggests that additional uses may be found for data, and 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 is a telling case in point, and provides a useful precedent for C.D.I.

Therefore, it may be argued that, as far as the quantitative data gathered for C.D.I. is concerned:

- there is no specific statement precluding archiving in the consent forms and information sheets;
- all of the quantitative data will be sufficiently anonymised, removing the data from the legal constraints of data protection;
- the various academic institutions' ethics committees have granted ethical approval for sharing the data;
- it is inappropriate to apply the ethical standards which would have applied in 2007 (when archiving was not generally considered in social science research studies) to the current data; and
- C.D.I. holds datasets which have a significant potential value to current and future researchers and that it might be ethically questionable to deny access to the datasets to bona fide researchers.

In a paper by Dr. Harry Comber under the title of 'Secondary use of data – striking a balance' to a conference organised by the Office of the Data

Protection Commissioner in November 2006, he says:

"However, against these undoubted benefits must be set the possibility of negative effects on the individual whose data is being used and, of course, the requirements of data protection law. At present, no framework exists in Ireland for deciding on the balance between the individual and public interest in data use and so the emphasis is currently on minimising the use of personal data. A number of methods are available for this, but all are based on data anonymisation." [10]

In addition, the Data Protection Act includes the following provision:

"6A.—(1) Subject to subsection (3) and unless otherwise provided by any enactment, an individual is entitled at any time, by notice in writing served on a data controller, to request him or her to cease within a reasonable time, or not to begin, processing or processing for a specified purpose or in a specified manner any personal data in respect of which he or she is the data subject if the processing falls within subsection (2) of this section on the ground that, for specified reasons—

> *"(a) the processing of those data or their processing for that purpose or in that manner is causing or likely to cause substantial damage or distress to him or her or to another person, and*

> *(b) the damage or distress is or would be unwarranted.*[11]*"*

In addition, a review of the literature on the subject in Ireland states that:

"Anonymisation may be described as 'the removal of name, address, full post code and any other detail or combination of details that might support identification'. Pseudonymisation of data differs from anonymised data as the original provider of the information may retain

a means of identifying individuals. Data that cannot identify an individual patient, either directly or through linkage with other data available to a user, do not need to be regarded as confidential." [12]

If this argument is accepted, then C.D.I. 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 traceable. If C.D.I. is satisfied that this is a slight risk, then it is possible to archive the quantitative datasets and some of the qualitative material.

Comber continues:

"Clearly, a universal requirement for consent to the use of personal health data is inconsistent with the most effective use of this data. At the same time, there is little evidence for any significant public concern about the use of health data for research or statistical purposes, or of any disadvantage accruing to data subjects."

In the case of the <u>qualitative</u> data, while it may be possible to anonymise at least some of the material, it may well be the case that the amount of work involved would outweigh the benefits of processing and archiving the material and, as Comber pointed out *"the price of the efforts at anonymisation may be serious degradation of the data."*

This Report proposes a plan which addresses the issues relating to archiving while maximising the potential benefits of the datasets to future researchers and to the greatest possible understanding of issues and strategies for childhood development. Retrospective consent can be sought from participants in cases where this is required. In practice, this may mean that

---

[10] H. Comber, Secondary use of data – striking a balance. Paper presented to the Conference Promoting health research & protecting patient rights. Office of the Data Protection Commissioner, November 2006. The authors are grateful to Dr. Suzanne Guerin for drawing this reference to their attention.

[11] http://www.irishstatutebook.ie/2003/en/act/pub/0006/print.html

[12] *Gavin, A., Kelly, C., Nic Gabhainn, S. and O'Callaghan, E. (2011) Key issues for consideration in the development of a data strategy: A review of the literature. Dublin: Department of Children and Youth Affairs. http://www.nuigalway.ie/hbsc/documents/lp__key_issues_data_strategy_nov_2011.pdf*

Report to The Childhood Development Initiative on archiving of C.D.I. data

only part or none of the primary qualitative data can be made available for secondary use. However, there may be process material which can be archived for use by other researchers.

Methodology

The authors of this report used the following methodology:

- Review of information in C.D.I. and meetings and correspondence with C.D.I. project staff
- Reading the published evaluation reports
- Interviews with the teams which carried out the evaluations and hold the data under discussion:
  - o *Dr Liam O'Hare, Dr Karen Kerr and Dr. Andy Biggart, Centre for Effective Education, School of Education, Queen's University Belfast*
  - o *Professor Catherine Comiskey, (Director, Centre for Practice and Healthcare Innovation and Professor of Healthcare Statistics) and Karin O'Sullivan, School of Nursing and Midwifery, Trinity College, Dublin*
  - o *Siobhan Keegan, Centre for Centre for Social & Educational Research, Dublin Institute of Technology*
- Telephone conversation with Dr John Canavan, National University of Ireland, Galway

- Discussions with
  - o *Julia Barrett, Research Services Manager, U.C.D. library, with responsibility for the Irish Social Sciences Data Archive;*
  - o *Dr Jane Gray, Irish Qualitative Data Archive, National University of Ireland, Maynooth;*
  - o *Dr. Suzanne Guerin, School of Psychology, University College Dublin and Vice-Chair, C.D.I.;*
  - o *Ms. Jan Stokes, Research Ethics Administrator, U.C.D.;*
  - o *Mark Ward, PhD candidate, School of Social Work and Social Policy, Trinity College, Dublin, Lecturer in Research Methodology, including SPSS, member School Ethics Committee; and*
  - o *Dr Matthew Woolard & Dr Libby Bishop, UK Data Archive.*
- A review of best practice internationally through desk research
- Consultation with industry specialists on best practice and practical issues in archiving this type of material
- Contact with standards bodies, particularly the Office of the Data Protection Commissioner.

B. Components of an archiving plan for C.D.I.

The studies undertaken, and the datasets created and used by C.D.I. and the evaluation teams, are of immense potential benefit to current and future researchers including, but not confined to, those already working on the projects. They can also contribute to effective policy making, improved professional practice and the wellbeing of Irish society and future generations of children.

Their contents should, therefore, be preserved and made available as far is this is consistent with ethical and legal constraints, particularly participants' implied or expressed consent.

The main questions arising are the extent to which retrospective consent from the participants is required, where to lodge the datasets and other material and how to ensure enduring usability and access for genuine researchers. This is discussed further in the recommendations.

These issues need to be resolved for both qualitative and quantitative data, as the potential secondary use of each type is very different in nature.

While there are significant legal constraints imposed by the data protection legislation, appropriately anonymised material, mostly quantitative data, can more easily be archived and made available than qualitative data.

It is therefore necessary to establish a system for storing, maintaining and making accessible as much as possible of the material generated in the course of the evaluations. This means, in effect, handing over control of all the material generated, or disposing of what cannot, or does not need to be preserved. This requires careful selection of a host institution and a detailed and comprehensive plan for the future of the material. Meticulous and very considered planning, management and execution of the plan are required.

Issues to be addressed

In this report, we discuss in more detail the following issues:

1. What type of material should be preserved?
2. How should the data for archiving or disposal be processed?
3. How should the data be anonymised?
4. How should the data be prepared for transfer?
5. What secondary data should be created?
6. Where should the archives be held?
7. How should access be managed?
8. How should relations with repositories be managed?
9. How should the use of the archive be promoted?

Where appropriate, the issues for qualitative and quantitative primary data, as well as particular issues for secondary and tertiary information, are considered separately under each heading.

1. What type of material should be preserved?

The following archiving plan provides for preservation of and continuing access to all of the datasets generated by the C.D.I., insofar as this is compatible with considerations of privacy and data protection discussed above.

This includes:

- primary material, the datasets and other documentation;
- secondary material, including the metadata, regardless of format; and
- tertiary material, or studies derived from the materials to be archived.

The primary material to be managed comprises qualitative, quantitative and process/ administrative information.

Those responsible for managing the archiving of these projects will need to review all relevant documentation, dispose of duplicates and conduct a risk assessment for all data.

For material which is to be retained, a timescale should be assigned for retention and/ or disposal. It is generally considered good practice to retain such materials for five years, but this may vary depending on the uses for which it is retained, the technical medium used to store and access it and the needs and capacity of the institution holding it. Some material is being used for continuing longitudinal studies and may be retained for longer. Where research teams are not permanent staff of the academic institutions, or where researchers leave the institutions or retire, great care must be taken to ensure that such material is actually disposed of after the 5 year period.

Secondary use dramatically increases the value for money spent on data collection and processing, and therefore providing datasets for use by bona fide researchers is an exemplary use of publicly or partially publicly-funded primary research: *"Using secondary data enables one to conduct studies of high-impact research questions with dramatically less time and resources than required for most studies involving primary data collection."*[13] Government policy is moving towards ensuring that data from publically funded research should be available for future use; in this context, C.D.I.'s datasets provide potentially useful material for future research.

Quantitative Data

The primary quantitative information includes datasets and the material from which they were developed, questionnaires, score sheets, etc. and associated ancillary and administrative material – e.g. manuals, code-books, interviewer instructions, consent forms, information sheets, ethical approval, Garda clearance, etc.

In terms of the quantitative data, the SPSS files can more easily be made available as long as they are clear, consistent, anonymised raw data. In the case of the evaluation team in Q.U.B., the material has already been prepared for transfer and the cost of anonymisation has been included in their contract with C.D.I. The material in T.C.D. will need more work but anonymisation and preparation of the material is considered manageable. The data in DIT will require considerable work to remove imputed data. Supporting material, i.e. metadata, needs to be provided for all of these, to ensure they can be located and appropriately exploited by the end user.

Qualitative Data

The primary qualitative data includes soundfiles, transcripts and reports of focus group and individual interviews, field notes, etc.

The Data Protection Commission's "Data Protection Guidelines on research in the Health Sector" states that: *"Irrevocable anonymisation of personal data puts it outside data protection requirements as the data can no longer be linked to an individual and therefore cannot be considered to be personal data."* [14]

Qualitative data presents different challenges; anonymising the data is exceptionally time consuming and will not be possible in all cases.

Issues of privacy and data protection are more sensitive for the qualitative data than for the quantitative data, since datasets are more easily anonymised provided the cell size is sufficiently large (i.e. 20 or over).

Anonymisation is critical to the ethical archiving of any research material but it presents problems which are discussed in Section 3 below.

The process material includes progress reports, minutes, newsletters and other administrative documentation.

[13] Smith, A.K. et al., Conducting High-Value Secondary Dataset Analysis: An Introductory Guide and Resources. J Gen Intern Med >v.26(8); Aug 2011. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3138974/ accessed 24 April 2013.

[14] See: http://www.dataprotection.ie/documents/guidance/Health_research.pdf

Metadata needs to be created for any material to be archived. This material will take 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 (as discussed further under section 5 below). 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.

The C.D.I. datasets were at least partly funded by the Department of Children and Youth Affairs and The Atlantic Philanthropies. A recent study published by the Department and funded by Atlantic Philanthropies on "Key issues for consideration in the development of a data strategy: A review of the literature" shows evidence of a commitment to developing a comprehensive data strategy for funded research.[15]

2. How should the data for archiving or disposal be processed?

Archiving primary data, whether qualitative or quantitative, is an evolving area and as such, while standards are gradually being drafted and adopted, it is inevitable that more comprehensive protocols will be developed and applied, in both the short and medium terms. Therefore any standards or requirements cited in this report are likely to change, perhaps with little warning. It is therefore perhaps an especially good time to archive data-sets, as more stringent requirements, involving more work, may be required in the future.

It will be necessary to consider all types of material and all individual files to ensure that any data sourced is suitably anonymised and appropriately formatted.

This will involve:

For all data:
- Listing the materials

For quantitative data:
- Removing duplicates and any 'imputed' data.
- Transferring files into appropriate file formats for archiving;
- Removing materials in formats not suitable for archiving after they have been transferred into appropriate formats.

For qualitative Data
- Removing duplicates;
- Assessing the risks attached to retaining documents and making them accessible;
- Removing material which cannot be anonymised and original material after an anonymised version has been produced;
- Transferring files into appropriate file formats for archiving;
- Removing materials in formats not suitable for archiving after they have been transferred into appropriate formats.

This is likely to generate a large amount of material which is not suitable for archiving; its disposal will need to be carefully managed.

In some cases, retrospective consent may be required, and may be relatively easily forthcoming. In general, however, it would be difficult to get retrospective consent from the parents and guardians. However, in line with the arguments above, this may not be necessary.

There is a number of options for such material, depending on the level of risk and the nature of the material.
- 'High risk' or 'non-anonymised' data may continue to be used by the evaluation teams which generated them and by the C.D.I. for use in academic publications. This will require close attention to secure storage and authorized access within the limits of data protection and the associated ethical guidelines;
- Material which cannot be appropriately anonymised because it comes from a small

15 Gavin, A., Kelly, C., Nic Gabhainn, S. and O'Callaghan, E. (2011): Key issues for consideration in the development of a data strategy: A review of the literature. Dublin: Department of Children and Youth Affairs. Available at: www.dcya.ie.

sample and could therefore be identified to a particular group or individual may be stored securely and embargoed for between 5 or 30 years, depending on the particular nature of the material and the purpose for which it was collected.

Decisions should be taken on a case by case basis following appropriate review.

All other material must be safely and securely disposed of, through thorough deletion of digital material and shredding of paper files.

Before destruction of material which cannot be suitably anonymised, or if it is to be embargoed, a full debriefing report should be written so that important lessons about the conduct of research projects may be learned by the community and to guide future researchers, unless C.D.I. is clear that the process evaluation has elicited the relevant information.

The length of time that particular sections of archives will be preserved or embargoed must be part of the formal agreement between the C.D.I. and the repositories. Best practice guides differ on this question. There has, in the recent past, been a view that *long-term* was loosely defined as 5 years, however, this is now under question. The precedent set by the Government and the CSO in releasing material after 30 years is followed in some archives.

The resources required depend on the current state of the datasets and the work that will be required to get them into a fit state for archiving. As this is not included in the original contracts of the research teams, they should be required to furnish estimates of the work needed, the costs, timescale and other requirements.

In the case of datasets such as those created by C.D.I., their preservation should be for an indefinite period provided that they are suitably anonymised.

The current state of progress is detailed in the Annex to this report.

3. How should the data be anonymised?

Sensitive data must be anonymised to the greatest extent possible. This is essential to ensuring that individuals or groups cannot be identified, as required by good practice and by data protection legislation.

Effective anonymisation is very time-consuming, especially for qualitative data. Guidelines for anonymising qualitative and quantitative are available on http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation?index=0

For quantitative data, the general rule of thumb that cell sizes of 20 or more ensures that data is sufficiently anonymised, will not be applicable in every case and depends on the materials in question. Examples of this would include cases where a small number of workers is involved in a project; a school Principal; public health nurses etc. Sound files, which can provide a rich source of information to future researchers, and may also be of interest to social-linguists, are particularly challenging in this regard.

It is also possible to over-anonymise 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[16]. 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.

Good practice in anonymising data of this type involves:
- Deletion of obvious identifying data (e.g. names, location, organisation);
- Removal of details which allow a person to be easily identified (address, occupation, etc.);
- Replacing these with descriptions congruent with the subject matter.[17]

[16] A. Clarke, Working paper: anonymising research data. 2006. ESRC Centre for Research Methods, NCRM Working Paper Series 7/06. http://eprints.ncrm.ac.uk/480/1/0706_anonymising_research_data.pdf Accessed 23 April 2013
[17] http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation

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 can lead us into a problematical area. This should be done in line with the guidelines from the Irish Social Science Data Archive of Quantitative Datasets and the Irish Qualitative Data Archive. In one academic institution which follows what is regarded as best practice, the 'tracking table' is retained in case it is needed for specific purposes in future, such as obtaining consent for the use of data, but is not accessible to researchers using the archives.

Where possible, this data should be anonymised by the teams who carried out the relevant evaluation teams subject to agreed guidelines and controls specified by C.D.I.

Next steps

1. For the teams who are clear that they have sufficient ethical approval to provide access to the quantitative data, the next stage is for them to estimate what would be required to prepare this material for archiving. Any dubiety about ethical approval should be clarified before moving to this step.

2. Informal discussions with the Irish Social Science Data Archive of Quantitative Datasets indicates that it is interested in C.D.I.'s datasets. Formal discussions should be opened with I.S.S.D.A. to:
   • establish whether they would be happy to provide access to the archive;
   • ascertain what their requirements would be;
   • agree how they will manage the primary and secondary access and;
   • agree the end-user licence agreement.

3. In relation to the qualitative material, it will be necessary to establish what, if any, of the material can be archived and what steps would be needed to arrange retrospective consent, particularly consent from C.D.I. staff for archiving material which originates from them;

4. Formal discussions should be opened with the Irish Qualitative Data Archive to:
   • establish whether they would be happy to provide access to the archive;
   • ascertain what their requirements would be;
   • agree how they will manage the primary and secondary access and
   • agree the end-user licence agreement.

5. A system must be established to create the metadata for all material for deposit.

Whilst the storage in separate archives makes some sense, the value of archiving the C.D.I. datasets as a whole should be considered to take precedence.

4. How should the data be prepared for transfer?

Data archives typically expect the data they receive to be technically and legally suitable for use, adequately documented and relatively error-free.

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

Time and other constraints did not allow a comprehensive review of all materials for this study. It will be necessary to systematically review all datasets to establish what type of preservation and access is appropriate and what processing is required, provided that retrospective consent has been obtained.

This is an enormous task, given the quantity of data available. In some cases, we are aware that attention has been paid to anonymising data stored by the teams, but this is by no means universal.

The scale of this 'clean up' is potentially enormous and has not been fully quantified. One evaluation team estimates that they hold more than a million individual pieces of data in their project alone. C.D.I. will clarify the willingness of the evaluation teams to ensure the quality of their documentation and its readiness for archiving.

Digital objects require pro-active intervention to remain accessible. Data on formerly standard 'floppy disks' can now only be retrieved with specialist hardware and the same fate may befall current data which is not properly stored. 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.

For these reasons, repositories specify acceptable formats; U.C.D. was clear that all material must be in non-proprietary file formats. Thus, for example, PDF should be used instead of Word for documents. This makes maintenance of the documentation easier, with reduced need for reverse engineering. U.C.D., for example, further insists that the same material should be provided in a number of different formats. However, N.U.I. Maynooth was more flexible on the technical requirements.[18] The U.K. Data Archive provides a list of format which are acceptable to it. (http://data-archive.ac.uk/create-manage/format/formats-table).

The current material is in a variety of formats. These include SPSS statistical datasets; RTF, PDF and Word documents, MP3 soundfiles, NVIVO, Mac PDA files and paper records. Preferred formats for qualitative files are specified at http://www.iqda.ie/content/deposit-data.

The appended table summarises the variety of formats.. Therefore, considerable work will be involved in re-formatting primary qualitative and quantitative data for deposit.

5. What secondary data should be created?

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.

Future researchers must be provided with sufficient secondary data (metadata) to enable them to locate and estimate the potential value of C.D.I.'s primary material for their research.

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.[19] Examples of such metadata may be found for any of the surveys in the Irish Social Science Data Archive[20], although the level of necessary detail will vary for each study.

A recent Canadian study notes that: "Each discipline has its own rules or customs for metadata, and most data repositories have formal metadata standards for submissions. If you plan to deposit your data in a repository, you will probably have to create metadata

---

[18] http://www.iqda.ie/content/deposit-data

[19] D.D.I. is the Data Documentation Initiative, a metadata specification for the social and behavioral sciences http://www.ddialliance.org/. For OAI-PMH see http://www.openarchives.org/pmh/ both sites accessed 3 April 2013.

[20] http://issda.U.C.D.ie/webview/

for your dataset or, at least, provide the repository with enough information to create metadata. Becoming familiar with the metadata standard in your discipline will make this process a lot easier. Metadata is often structured as a series of fields and recorded in XML format."[21]

6. Where should the archives be held?

In considering where to lodge the datasets and other material and how to ensure retention and long-term preservation and access for accredited researchers, consideration must be given to digital capacity, I.T. skills, trustworthiness, and the ethical approval and data protection legislation. It is recognised that it is difficult if not impossible to anonymise a significant proportion of the material, so any plan must create adequate safeguards to ensure protection of project participants and the best possible use of the data.

As C.D.I. does not have and is unlikely to acquire the technical infrastructure, space, expertise, finance and staff resources to manage these archives, they must be deposited in one or more appropriate institutions to ensure long-term access and viability.

All material to be archived must be appropriately stored, managed, preserved and accessed. The cleaned-up, anonymised master-files must be archived in an appropriate location. Research teams may continue to hold and use their current files for the purpose of their research. Agreement should be reached with the evaluation teams' institutions about the longevity and access conditions attaching to these files and their eventual disposal or deposit.

It is always difficult to predict the future research landscape with any certainty. However, technical architecture, file formats and user profiles and expectations will all change. The holder of the documentation must therefore demonstrate a commitment and a capacity to maintain the usability of the material, regardless of format; this implies

significant technical capacity and skills, as well as excellent administrative systems and structures.

Future users will include post-graduate students, established researchers, policy makers, social practitioners, and others, possibly including survey participants and their families. Robust systems, controls and appropriate end-user agreements must be in place to facilitate effective use of the data, protect the data from misuse and ensure appropriate access to those who need this data. (These are outlined in Section 8).

The material must be archived in a reputable institution with a secure future and technical competence, knowledgeable staff and adequate infrastructure.

The archiving, preservation and control of the datasets requires a team with a thorough understanding of good practice and ethical issues in digital archiving and the appropriate technical and administrative skills, who can be relied on to maintain and update the storage media and manage access.

The institutions should also be accredited, or at least working towards accreditation.

As it is desirable that the archives should be located in Ireland, the obvious options for deposit are U.C.D., which hosts the Irish Social Science Data Archive of Quantitative Datasets[22], for the quantitative data, and N.U.I. Maynooth, which hosts the Irish Qualitative Data Archive.[23]

It is regrettable that it would be necessary to split the archives and it would therefore be very important that appropriate linkages be established and maintained to make potential users aware of the existence and contents of both archives.

U.C.D. is currently working towards accreditation under CESSDA (Council of European Social Science

---

[21] Research Data Repositories by Steve Marks Feb 25, 2013 on the Scholars Portal, Ontario Council of University Libraries, http://guides.scholarsportal.info/researchdata

[22] http://www.U.C.D.ie/issda/.

[23] http://www.iqda.ie/

Data Archives)[24] and is linked in this process to the UK Data Archive[25], one of the leaders in its field.

Negotiations should therefore be initiated with U.C.D. and NUI Maynooth. This will require, for each institution, the development of a Submission Information Package (SIPS); an Archival Information Package (AIP) and a Dissemination Information Package (DIP) which will specify all relevant standards and processes for deposit, storage, preservation and dissemination.

The research evaluation teams should be required to lodge copies of their published reports in their parent institutions' repositories and in the Irish Health Repository and www.lenus.ie. As Lenus's contents are included in the national aggregator for institutional repositories http://rian.ie/ C.D.I.'s research will be easily retrievable worldwide simply by lodging the reports in this digital repository, at no additional cost to C.D.I.

7. How should access be managed?

As part of the agreement for the dissemination information package, C.D.I. must reach agreement with the hosting archives about levels of access, access procedures and end-user licensing agreements. In the case of the quantitative data, we recommend that single level access is appropriate. In the case of the qualitative data, some users may request secondary access to the original data and appropriate procedures must be put in place to guide and manage this.

Access should be permitted by completion of an application form to the data archives. An example may be downloaded from http://www.ucd.ie/issda/.

All users should be required to sign an end user agreement outlining how and for what purpose the data can be used. This must take into account the risk associated with each type of document.[26]

The end-user licencing agreement should also include a provision that users lodge a copy of the published results of their research in their own organisations' repositories, inform the data archive depository and supply a copy to U.C.D. and N.U.I.M. for the quantitative and qualitative data respectively. Where they have no institutional affiliations they should submit copy of their published research to an agreed repository.

The agreement should also specify how the datasets are to be cited. This is important not only for accountability and for the convenience of future users but to promote awareness and use of the archives.

These provisions should be enforced particularly in the case of any studies subsidised by bursaries associated with C.D.I. or the archives.

8. How should relations with repositories be managed?

C.D.I.'s contracts for most of the evaluations, state:

> 'Any Intellectual Property developed, acquired, made or discovered by [the research team] during the course of this contract with C.D.I. in connection with or in any way affecting or relating to the business of C.D.I. shall belong to and be the absolute property of C.D.I..[The research team] hereby assigns to C.D.I. all rights in such Intellectual Property for their full term throughout the world. C.D.I. hereby grant a royalty free, non-transferable, non exclusive licence to [the research team] and their agents to utilise the intellectual property rights subject to the necessary protections which may arise under relevant legislation, including Data Protection Legislation. In addition, the parties hereby grant a royalty free, non-transferable, non-exclusive licence to The Department of Children and Youth Affairs, The Atlantic Philanthropies and their agents to utilise the intellectual property rights subject to the

---

24 http://www.cessda.org/

25 http://data-archive.ac.uk/

26 Relevant examples of end-user agreements can be found at http://data-archive.ac.uk/media/381244/ukda137-enduserlicence.pdf, http://www.iqda.ie/sites/default/files/IQDA_Data_Access_Request_Form.pdf and via a link from http://www.U.C.D.ie/issda/

necessary protections which may arise under relevant legislation, including Data Protection Legislation.'

Given the uncertainty over the future and role of the C.D.I., it will not be possible to enforce this directly. It will also be impossible for researchers to ask for information or additional access, as is provided by other data-set publishers such as the C.S.O.

Provision governing this must be included in the dissemination information packages agreed with the host archives.

C.D.I. and the host repository must agree:

- **A Preservation Implementation Plan:** A written statement, authorised by the management of the repository, which describes the services to be offered by the repository for preserving objects accessioned into the repository in accordance with the Preservation Policy;
- **A Preservation Policy:** A Written statement, authorised by the repository management, that describes the approach to be taken by the repository for the preservation of objects accessioned into the repository. The Preservation Policy is consistent with the Preservation Strategic Plan;
- **A Preservation Strategic Plan:** A written statement, authorised by the management of the repository, stating the goals and objectives for achieving that part of the mission of the repository concerned with preservation. Preservation Strategic Plans may include long-term and short-term plans.
- **An Access Policy:** A Written statement, authorised by the repository management, that describes the approach to be taken by the repository for providing access to objects accessioned into the repository. The Access Policy may distinguish between different types of access rights, for example between system

administrators, Designated Communities, and general users.[27]

9. How should the use of the archive be promoted?

As already indicated, it is to be hoped that the user population for the material discussed in this report will broaden and deepen as time goes by. This is in line with experience of other similar archives in the literature.

To be of broad benefit to Irish children and Irish society, the material 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 which use the C.D.I. data archive 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' institutional 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.[28]

In addition, the datasets should be listed in Thompson Reuters Data Citation Index, as U.C.D. would do as a matter of practice if they act as one of the repositories.[29]

It is recommended that a blog be started by C.D.I., and subsequently handed over to an appropriate institution, to promote awareness of the archives and the archiving project. See http://www.iqda.ie/blogs/irish-qualitative-data-archive.

---

[27] International Organisation for Standardization. (2012). Space data and information transfer systems: audit and certification of trustworthy digital repositories = Systèmes de transfert des informations et données spatiales : audit et certification des référentiels numériques de confiance. Geneva, Switzerland, International Organization for Standardization. http://public.ccsds.org/publications/archive/652x0m1.pdf (accessed 2 April 2013)

[28] http://www.eric.ed.gov

[29] http://wokinfo.com/products_tools/multidisciplinary/dci/ accessed 3 April 2013.

In addition, 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.)
- C.D.I.'s Twitter feed
- Presentations and poster sessions at conferences, research seminars etc.

C. Recommendations

A. Overall Recommendations

In all future research commissioned by C.D.I., the research teams should consider the need to archive data and the requirements to do so and these considerations should be built into the research proposals from the start.

Support should be provided to the research community in relation to archiving as required.

A toolkit should be developed to guide the archiving of data from current and future research studies.

B. General Recommendations for archiving of existing materials commissioned by C.D.I.

C.D.I. should seek to deposit its <u>quantitative</u> data with the Irish Social Science Data Archive in U.C.D., and its <u>qualitative</u> data in the Irish Qualitative Data Archive in N.U.I. Maynooth. Contingent upon this, their own guidelines should be used in drawing up all plans and agreements.

All data to be transferred must be suitably anonymised and migrated to appropriate formats as specified by the relevant repository.

Appropriate linkages between the qualitative and quantitative materials should be created and all agreements with repositories should ensure that these linkages are maintained.

Comprehensive, accessible and clear metadata should be created for all datasets and documents.

C.D.I. and the receiving institutions should agree the following, as described in Section 8 above:

- A preservation plan
- A preservation policy
- A preservation strategic plan
- An access policy and
- Agreed templates for access agreements and end-user agreements and accompanying guidelines.

Agreements should be reached with the research teams to store or destroy all the data from the projects over a phased basis, allowing for the accepted period of 5 years retention, bearing in mind their own needs for future research on the data.
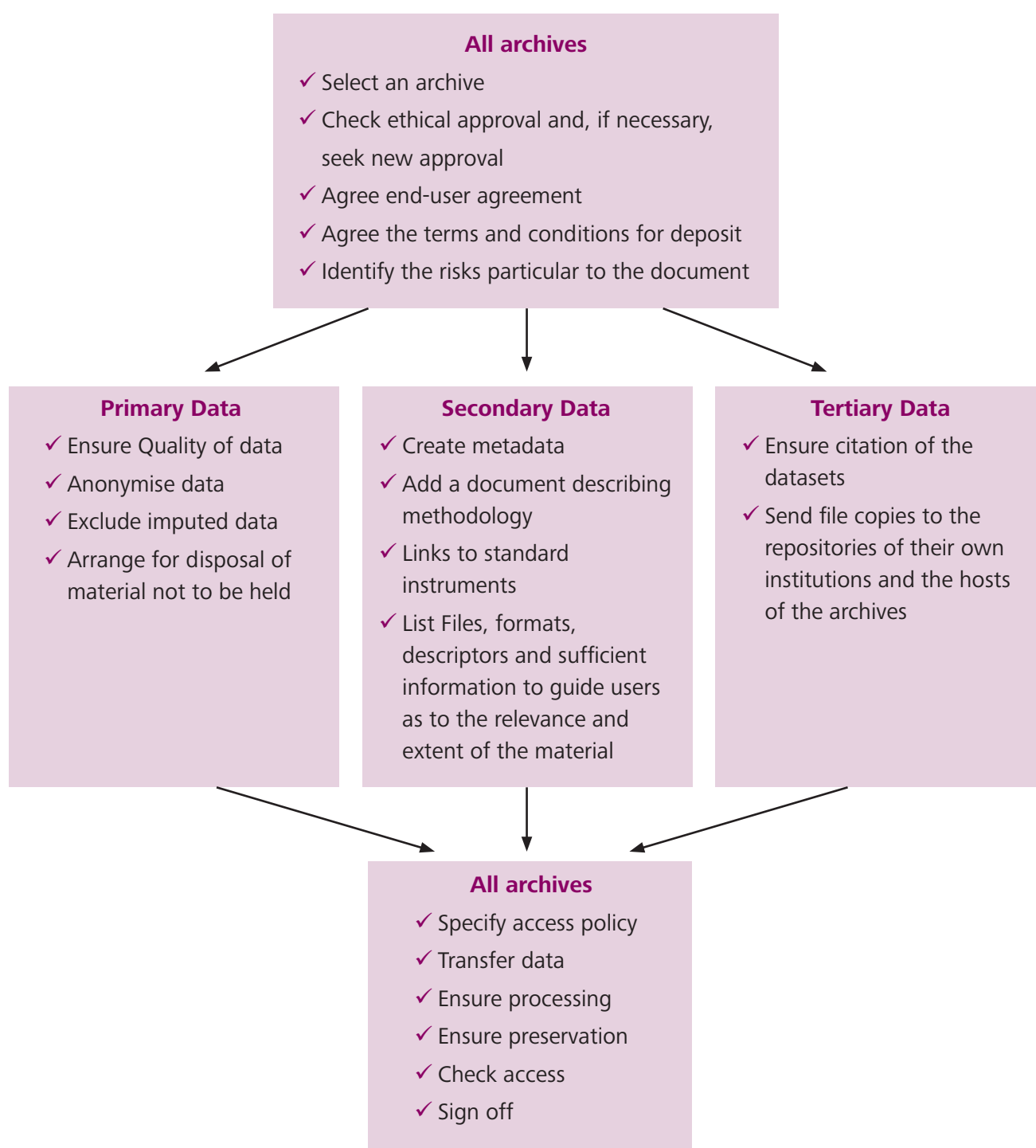
The possibility of archiving and secondary use should be included in planning any future significant research which is likely to lead to the creation of valuable data-sets or other important research data. This requires including a provision for archiving in all consent forms, information sheets and ethical approval requests.

The next steps are as follows:

1. The evaluation teams should estimate what would be required to prepare this material for archiving. The situation in respect of ethical approval should be clarified before moving to this step where this has not already been done.
2. Formal discussions should be opened with the Irish Social Science Data Archive of Quantitative Datasets to:
   - establish whether they would be happy to provide access the archive
   - ascertain what their requirements would be
   - agree how they will manage the primary and secondary access and
   - agree the end-user licence agreement.
3. In relation to the qualitative material establish what, if any, of the material can be archived and what steps would be needed to arrange retrospective consent?
4. Formal discussions should be opened with the Irish Qualitative Data Archive to:
   - establish whether they would be happy to provide access to the archive
   - ascertain what their requirements would be
   - agree how they will manage the primary and secondary access and
   - agree the end-user licence agreement.
5. Systems should be established to create metadata for all the material to be lodged.

6. C.D.I. should create a blog to advertise the existence of the datasets and promote their use by bona fide researchers.

7. Before destruction of material which cannot be suitably anonymised, or is to be embargoed, a full debriefing report should be written so that important lessons about the conduct of research projects may be learned by the community and to guide future researchers, unless C.D.I. is clear that the process evaluation has elicited the relevant information.

The proposed methodology for transferring files is laid out in the table below. Specific recommendations to C.D.I. in relation to each of the teams are also attached.

Proposed methodology for processing archives

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

Primary Data
- ✓ Ensure Quality of data
- ✓ Anonymise data
- ✓ Exclude imputed data
- ✓ Arrange for disposal of material not to be held

Secondary Data
- ✓ Create metadata
- ✓ Add a document describing methodology
- ✓ Links to standard instruments
- ✓ List Files, formats, descriptors and sufficient information to guide users as to the relevance and extent of the material

Tertiary Data
- ✓ Ensure citation of the datasets
- ✓ Send file copies to the repositories of their own institutions and the hosts of the archives

All archives
- ✓ Specify access policy
- ✓ Transfer data
- ✓ Ensure processing
- ✓ Ensure preservation
- ✓ Check access
- ✓ Sign off

C. Recommendations in respect of material developed by the team at Trinity College, Dublin

1. C.D.I. or the team should clarify the ethical approval required to archive some or all of the data

2. If it is possible to archive some or all of the data, a plan should be drawn up for its transfer along the lines indicated in this report and in the toolkit to be developed

3. Discussions with the nominated repositories on their requirements for the transfer of the data and a practical agreement on this transfer should be commenced

4. In the case of any material which cannot be archived, the conditions for storage and disposal should be agreed between C.D.I and the team.

5. Plans for use of this data in future research should be agreed.

D. Recommendations in respect of material developed by the team at Dublin Institute of Technology

1. The team should clarify what is to be done to archive this material

2. An estimate of the resources and funds required to carry out this work should be made and discussed with C.D.I.

3. Discussions with the nominated repositories on their requirements for the transfer of the data and a practical agreement on this transfer should be commenced

4. In the case of any material which cannot be archived, the conditions for storage and disposal should be agreed between C.D.I and the team.

5. Plans for use of this data in future research should be agreed.

E. Recommendations in respect of material developed by the team at Queens University, Belfast

1. The team should clarify what is to be done to archive this material

2. An estimate of the resources and funds required to carry out this work should be made and discussed with C.D.I.

3. Discussions with the nominated repositories on their requirements for the transfer of the data and a practical agreement on this transfer should be commenced

4. In the case of any material which cannot be archived, the conditions for storage and disposal should be agreed between C.D.I and the team.

5. Plans for use of this data in future research should be agreed.

References

ALA (American Library Association) (2007). *Definitions of digital preservation.* Chicago: American Library Association. Available at: http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.pdf

A. Clarke, Working paper: *anonymising research data.* 2006. ESRC Centre for Research Methods, NCRM Working Paper Series 7/06. http://eprints.ncrm.ac.uk/480/1/0706_anonymising_research_data.pdf Accessed 23 April 2013

H. Comber, Secondary use of data – striking a balance. Paper presented to the Conference Promoting health research & protecting patient rights. Office of the Data Protection Commissioner, November 2006. The authors are grateful to Dr. Suzanne Guerin for drawing this reference to their attention.

Cragin, Melissa H., et al. "Data sharing, small science and institutional repositories." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368.1926 (2010): 4023-4038. http://rsta.royalsocietypublishing.org/content/368/1926/4023.full

Data Protection (Amendment) Act 2003, No. 6 of 2003 (see http://www.irishstatutebook.ie/2003/en/act/pub/0006/print.html)

Gavin, A., Kelly, C., Nic Gabhainn, S. and O'Callaghan, E. (2011) *Key issues for consideration in the development of a data strategy: A review of the literature.* Dublin: Department of Children and Youth Affairs. http://www.nuigalway.ie/hbsc/documents/lp__key_issues_data_strategy_nov_2011.pdf

Hrynaszkiewicz, Iain, and Douglas G. Altman. "Towards agreement on best practice for publishing raw clinical trial data." *Trials* 10.1 (2009): 17. http://www.biomedcentral.com/content/pdf/1745-6215-10-17.pdf accessed 7 April 2013,

Irish Qualitative Data Archive and Tallaght West Childhood Development Initiative, *Best Practice in Archiving Qualitative Data.* (2010). http://research.nuim.ie/sites/ research.nuim.ie/files/images /IQDA%20Best%20Practice%20Handbook.pdf

International Organisation for Standardisation. (2012). *Space data and information transfer systems: audit and certification of trustworthy digital repositories = Systèmes de transfert des informations et données spatiales: audit et certification des référentiels numériques de confiance.* Geneva, Switzerland, International Organization for Standardisation. http://public.ccsds.org/publications/archive/652x0m1.pdf (accessed 2 April 2013)

Irish Research Council Government Of Ireland, Research Project Grants Scheme, 2013, Terms And Conditions, http://www.research.ie/sites/default/files/irc_rpg2013_terms__conditions_final_0.pdf

King, Gary: Ensuring the Data-Rich Future of the Social Sciences *Science* 331, 719 (2011); http://bidm.stat.fju.edu.tw:81/2013--研究所Data%20Mining/DM2013-0305/science-2011-2-11-/Science-12.pdf

Marks, SteveResearch *Data Repositories*, Feb 25, 2013 on the Scholars Portal, Ontario Council of University Libraries, http://guides.scholarsportal.info/researchdata

NISO: Dublin Core Metadata Element Set Approved. Source: http://www.niso.org/news/releases/PRDubCr.html (accessed on 14th June 2008). Quoted in Jiban K. Pal, Practical framework for harvesting standard metadata in digital repository. 7th International CALIBER-2009, Pondicherry University, Puducherry, February 25-27, 2009 © INFLIBNET Centre, Ahmedabad http://www.inflibnet.ac.in/caliber2009/CaliberPDF/14.pdf.

OECD Glossary of Statistical Terms - Data imputation Definition stats.oecd.org/glossary/ detail.asp?ID=3406

Palmer, Carole L., Nicholas M. Weber, and Melissa H. Cragin. "Analytic potential of data: assessing reuse value." *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries.* ACM, 2011.

Social Science Data Archive Policy on Acquisitions and Archiving (April 2011). http://dataarchives.ss.ucla.edu/Policy%20on%20Archival%20Operations.doc

Smith, A.K. et al., Conducting High-Value Secondary Dataset Analysis: An Introductory Guide and Resources. J Gen Intern Med >v.26(8); Aug 2011. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3138974/ accessed 24 April 2013.

SWAN, A., & BROWN, S. (2008). To share or not to share publication and quality assurance of research data outputs; main report. London, Research Information Network]. http://rinarchive.jisc-collections.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs

Tenopir, Carol, et al. "Sharing data: Practices, barriers, and incentives." *Proceedings of the American Society for Information Science and Technology* 48.1 (2011): 1-4.

Uhlir, Paul F., and Peter Schröder. "Open data for global science." *Data Science Journal 6* (2007): OD36-OD53.

U.K. Data Archive. *Across the decades: 40 years of data archiving.*(2007).http://data-archive.ac.uk/media/54761/ukda-40thanniversary.pdf accessed 15 August 2013.

Thomson, Denise; Bzdel, Lana; Golden-Biddle, Karen; Reay, Trish & Estabrooks, Carole A. (2005). Central Questions of Anonymisation: A Case Study of Secondary Use of Qualitative Data [33 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research,* 6(1), Art. 29, http://nbn-resolving.de/urn:nbn:de:0114-fqs0501297.

Websites

http://www.digitalpreservationeurope.eu/what-is-digital-preservation/#96, (accessed 28th. March 2013.)

http://trusteddigitalrepository.eu/Site/Welcome.html

http://www.ddialliance.org/

http://www.openarchives.org/pmh/

http://www.U.C.D..ie/issda/

http://www.iqda.ie/

http://www.cessda.org/

http://data-archive.ac.uk/

http://prijipati.library.usyd.edu.au/bitstream/2123/2676/1/LegalFramework_Ch8.pdf

http://www.eric.ed.gov

http://wokinfo.com/products_tools/multidisciplinary/dci/, accessed 3 April 2013

Annex – List of Projects and Position in relation to archiving

List of projects

Doodle Den, Mate-Tricks; Centre for Effective Education; Queens University Belfast

Projects	Location	Quantitative data	Qualitative data
Doodle Den, Mate-Tricks	Centre for Effective Education; Queens University Belfast	Data is close to being cleaned up, anonymised and ready to be lodged	
Speech and Language Therapy Model, Early Childhood Care and Education	Centre for Social and Education Research; Dublin Institute of Technology	Data can be cleaned up, anonymised and lodged but with more work than Q.U.B.	
Healthy Schools Programme	School of Nursing and Midwifery; Trinity College Dublin	Ethical agreement needs to be clarified before material can be prepared for archiving	
Community Safety Initiative, Restorative Practice Evaluation and the overall Process Evaluation	Child & Family Research Centre, School of Political Science & Sociology; National University of Ireland, Galway	Material will be retained for five years and destroyed	Material will be retained for five years and destroyed

Indicative list of typical file formats and proposed actions

Doodle Den, Mate-Tricks; Centre for Effective Education; Queens University Belfast

	Paper	Digital	Notes	Proposed action
Observations of Sessions	Paper files	Word, RTF, SPSS, Mac PDA		Embargo sound files. Archive SPSS
Focus Groups with Parents	Paper files	Word, RTF, SPSS, MP3, Mac PDA		Embargo sound files. Archive SPSS
Interviews with facilitators	Paper files	Word, RTF, SPSS, MP3, Mac PDA	Hard to anonymise	Retain 5 years and destroy
Interviews with teachers	Paper files	Word, RTF, SPSS, MP3, Mac PDA	Hard to anonymise	Retain 5 years and destroy
Interviews with principals	Paper files	Word, RTF, SPSS, MP3, Mac PDA	Hard to anonymise	Retain 5 years and destroy

Healthy Schools Programme, School of Nursing and Midwifery; Trinity College Dublin

	Paper	Digital	Notes	Proposed action
Qualitative data – 1-1 interviews		Sound files, also transcribed		
Teacher survey (S.P.S.S.)		SPSS files		
Videos		Video format	Not clear if they survived	
Manuals of practice	Printout	Word, RTF	Of interest to researchers.	Send to Google Books?
Ethical approval forms	Printout	Word, RTF		Retain 5 years and destroy
Garda Vetting information	Paper files			To be destroyed
Absenteeism, Immunisation, dental, BMI	Paper files	SPSS files	Very identifiable	Retain 5 years and destroy
Survey monkey results		SPSS files		
Admin: minutes, progress reports		word / pdf		

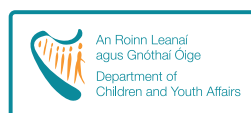## Speech and Language Therapy Model, Centre for Social and Education Research; Dublin Institute of Technology

	Paper	Digital	Notes	Proposed action
Parent focus groups	Paper files	MP3	No consent to share with C.D.I.	Retain 5 years and destroy
Speech and language therapists	Paper files	MP3	No consent forms	Retain 5 years and destroy
Interviews with 1 manager, 1 manager/ facilitator, 1 trainer	Paper files	MP3	very identifiable.	Retain 5 years and destroy

Early Childhood Care and Education, Centre for Social and Education Research; Dublin Institute of Technology

	Paper	Digital	Notes	Proposed action
Focus groups of parents and focus groups of staff	Paper files	Sound (windows or MP3?)	No one managing. Keeping database in case of funding for longditudinal study.	Keep 5 years then dispose
Assessments (British Ability Scales for core skills and fine motor skills	Paper (coded by ID Nos.)	SPSS anonymised at child level, not group level		Retain 5 years and destroy
Assessments – environmental scales (ECERS scale)	(12 to 15 file boxes)		Need time to sort Want to keep on DIT server	Qualitative can be archived Quantitative hard to anonymise
Qualitative data	NVIVO files			Not to be archived