Data Archive and Management Policy

This document aims to support the Data Manager in clarifying and standardising DCAL data practices

Last updated: December 2016, CY

Contents

1. Introduction and Context

DCAL is committed to ensuring that the outputs of the research centre, including research data, are managed and used in ways that will benefit further research, researchers, and the public (where possible).

The electronic archiving of research data generated by DCAL, the leading research centre on deafness and sign language in Europe, will play an important role in the creation of further scholarly content relating to d/Deafness studies internationally. DCAL will ensure that the use and reuse of research data is managed through consideration of relevant data ethics issues, and the centre will follow all requirements set out in the Data Protection Act 1998.

Making research data widely available to the research community in a timely and responsible manner ensures that these data can be verified, built upon and used to advance knowledge and for public engagement. In order to maximise the value of research data, DCAL (as both a data generator and data user) will act with integrity and transparency in managing, using, archiving, and sharing research data.

This Data Archiving and Management Plan outlines our goals to preserve and share research datasets in a manner that maximises their long-term ethical research use and value. The plan reflects ESRC and UCL best research practise guidelines (as well as best practice guidelines from the ESRC, DCAL's funder from 2006 to 2015), ensuring that research data is effectively and appropriately handled, organised, documented, and enhanced.

2. Data Policy Statement

a. A Guide for Researchers

i.   DCAL expects all researchers (both DCAL-funded and DCAL-associated) to maximise the availability of research data with as few restrictions as possible. This should be considered in the planning and proposals stages of research, and appropriate advice must be sought when constructing participant research (Appendix I) and video (Appendix II) consent forms, project information sheets (Appendix III), and data management plans e.g. as part of funding proposals (see Appendix IV). This involves thinking critically about how research data can be shared, how data can be anonymised, what might limit or prohibit data sharing (data ethics and protection practises, for example), and whether any steps can be taken to remove such limitations.

ii.   In cases where generated data will hold sensitive or personal information, participants and respondents must be made aware that their data will be stored securely by DCAL in line with the Data Protection Act 1998, and permissions for further research use must be sought.

iii.   Where the proposed research is likely to generate data outputs that will hold significant value as a resource for the wider research community, researchers will be required to submit a data management plan (e.g. the Data Management Plan submitted as part of the funding application or see template in Appendix IV) to the DCAL Data Manager so that arrangements can be made for future archiving and storage, including information relating to how data will be prepared, labelled, organised, and submitted for archiving (see the UCL Data Retention Schedule for more information).

iv.   When preparing research data for archiving, researchers must present and label their information in such a way that researchers outside of DCAL can clearly and basically understand the contents of the data sets without having to make contact with us; this can be made easier with additional explanatory documents such as Parts A and B of the Project Details form (see point 3.a.viii. and Appendices VI and VII below).

v.   All digitised data must be given to the Data Manager at the end of the project; they will then upload it to the Research Data Archive.

vi.   DCAL will:
- review data management and sharing plans, and any costs involved,
- work with researchers on an ongoing basis to support them in maximising the long-term value of key datasets resulting from their research,
- foster an environment that enables researchers to maximise the value of research data, and ensure that key data is archived and managed responsibly (as set out in the DCAL Data Archiving and Digitisation Plan).

vii.   DCAL expects all users of research data to acknowledge the sources of their data and to abide by the terms and conditions under which they accessed the original data.

3. Archiving Overview

DCAL has a rich collection of research data which has been collected since 2005. For the purpose of meeting funder and grant requirements and regulations, relevant materials will be digitised and archived. The centre's research, funded by the ESRC, must be deposited with a responsible repository as a contractual requirement at the end of DCAL. Where time and resources allow, DCAL-associated projects will also be digitised and archived.

This document (and particularly sections 3 and 4) attempts to formalise DCAL's approach to the digitisation and archiving of research data and aims to address the entire lifecycle of digitised content, from selection for digitisation, to the long-term curation of digitised content.

DCAL will work with UCL Library Services to create its own 'responsible repository' – an online Research Data Archive (for existing digital collections at UCL, see http://www.ucl.ac.uk/library/digital-collections). With regard to the principles set out in this policy and the Data Protection Act 1998, the DCAL Research Data Archives which will serve to (1) store all research data and content relating to DCAL-funded projects (thus meeting ESRC submission requirements), (2) curate and archive research data and content, and make it available securely online at least to DCAL PIs and if possible other DCAL-funded researchers, (3) make research data and content available to researchers outside of DCAL for the sole purpose of further research, where copyright and ethics (consent) permit.

DCAL partners will include UCL Library Services, and others will be considered if necessary.

a. Selection and Preparation of Content for Archiving

i.   Content will be prioritised for archiving following consideration of funder requirements and deadlines. Initially, all DCAL-funded work will be digitised and archived as per requirements of ESRC funding.

ii.   Where time and resources allow, data and information from DCAL-associated programmes will also be digitised and archived. Content will be prioritised for digitisation from internal resources where there is active research being carried out within DCAL that will benefit and contribute to further research. Consultation will take place via DCAL Directors Meetings and regular contact with PIs of DCAL-associated projects, as determined by an annual DCAL Data Holdings Survey.

iii.   All data linked to research projects must be digitised. Physical data including but not limited to paper, VHS, DVDs, etc. must be digitised and uploaded to the DCAL Research Data Archive. Researchers are responsible for the digitisation of their data, but the Data Manager will be available to support this endeavour. Each format will have its own technical requirements.

iv.   Filled-in consent forms must be digitised and stored on the Research Data Archive (only the PI will have access to this). The original hard copies will be securely recycled once the digital versions are securely archived.

v.   When creating digitisation plans, it is necessary to consider what will happen to 'original' data (i.e. paper, tapes, etc.) and if and how this will be stored and conserved. In these digitisation plans, DCAL must state whether a master copy of archived data will be kept, by whom, where, and how it will be (regularly) managed.

vi.   All personal data (including video data showing participants' faces) must be archived with the appropriate access restrictions and contact information relating to how access permissions can be sought.

vii.   When compiling data for digitisation and archiving, researchers (whether they have workspace at DCAL or not) must state if they have any data, or know of any data, stored at 49 Gordon Square. **Data that is not claimed may be destroyed, particularly if it holds personal or other sensitive information that cannot be associated to any project.**

viii.   It is the responsibility of the researcher to prepare and organise their data for digitisation and archiving. The cataloguing of data and metadata must be done in such a way that researchers outside DCAL are able to clearly understand at a basic level what data is available under each project title without having to contact the centre for further clarification. See Appendices VI and VII for DCAL's guidelines on data organisation (including plans for folder structure, labelling, associated 'Read Me' files, etc.); these can be used when preparing data. The Data Manager will be available to support DCAL researchers with any queries relating to this.

ix.   Researchers must attach metadata to their data and data sets; DCAL follows IMDI standards (see 4.v. and Appendix V) for this.

x.   The Data Manager will be available to support both DCAL and outside researchers with any queries relating to the DCAL Research Data Archive (see Contact information).

b. The Digitisation Process

i.   The physical process of digitisation and conversion to appropriate formats should be carried out using the most cost-effective method, depending on the scale and requirements of each project. Where possible in-house facilities will be used.

ii.   The choice of format will in part be influenced by recommendations from UCL Library Services and with consideration of the Research Data Archive's storage and archiving requirements. There should be no need to digitise any item a second time for a different purpose.

iii.   Following UCL Library Services guidelines, the quality of certain data (for example video data) must be maintained where possible. Any compression of data must be carried out to a high or good quality, and the potential for the degradation and bit rot of certain files and formats must be considered. Recommended formats and settings for video will be provided by the Data Manager in relevant instances.

iv.   Appropriate administrative metadata, including technical, rights, preservation and structural metadata will be created as part of the digitisation process and associated with the digital objects, to support both access and preservation. All such metadata will be standards-based. In general, metadata will loosely follow DCAL's IMDI standards (see Appendix V).

4. DCAL Research Data Archive

The DCAL Research Data Archive (created through UCL Library Services) will preserve quality, assure digital preservation, and conversion/ upgrades of research data where necessary.

a. Presenting Digitised Content Online

i.   Digitised content online should be suitable for the widest possible range of potential uses by different groups for different purposes. Due to the 'research' nature of DCAL data, the use and reuse of the centre's online materials will be limited to research use only (as outlined in the Data Protection Act 1998).

ii.   Where there are ethical limitations regarding the use and reuse of certain data, appropriate access restrictions will be imposed. DCAL Research Data Archive will attribute three levels of access to data: (1) open access and available to the public; this can be done with quantitative data (which has been anonymised) without explicit participant permission, (2) limited to researchers (who can contact the DCAL Data Manager for permission to access it), (3) restricted for use only within DCAL, and (4) locked and only accessible by project PIs, i.e. where consent forms and child data are highly restricted.

iii.   Anonymised research made publicly available on an open access basis will require that end users agree to use information for research only; this will be implemented through a terms and conditions acknowledgement which users must comply to.

iv.   High quality summaries and descriptions of data sets will be attached to each digital collection; this will assist users in understanding the quality of the DCAL research data. Where only part of a collection has been selected for digitisation or archiving, that should be made clear to users and information provided alongside the digitised content about the rest of the collection, and about related collections.

v.   Descriptive and demographic metadata – conforming to DCAL's IMDI standards (see Appendix V) - should be associated with each digitised item, wherever possible, or at least to subsets of different types of data. All digitised metadata will be made as visible as

possible to external search engines, the UK Data Archive, other research portals, and specialist aggregators; this will serve in making DCAL data searchable and more accessible.

vi.   Where technically possible and where resources permit, optical character recognition (OCR) should be used to make digitised text searchable; there are online OCR tools available which the Data Manager can help researchers use. Transcripts or contextual information might be needed to make the content more useable.

vii.   For researchers including digitised and electronic content in the digital collection: All digitised content online should include a clear statement of copyright, including end-user rights, and relevant terms and conditions.

viii.   Regular evaluations will be carried out of DCAL's archived content. Usage statistics, occasional user surveys, and online forms will be used to measure the amount and type of usage, and the degree of user satisfaction. That information will inform future decisions about archiving of new collections and the maintenance of existing archived collections.

ix.   DCAL's Research Data Archive will be presented online with the assistance and direction of UCL Library Services. The layout or presentation of online content will be reviewed by the Data Manager annually to ensure that information is clearly viewable and accessible.

5. Management

a. Ethics and Data Protection

DCAL is committed to using and generating research following best practise guidelines as set out by the UCL Research Ethics Committee. Please see the UCL REC website (https://ethics.grad.ucl.ac.uk/procedures.php) for information relating to ethics and data protection as well as ethics information for the Division of Psychology and Language Sciences (http://www.ucl.ac.uk/pals/intranet/ethics).

If you are storing any personal data about your participants, you need to obtain a **Data Protection Registration Number** by downloading and completing the UCL Research Registration Form (Form 2) which will need to be sent to data-protection@ucl.ac.uk**. The UCL Data Protection team will provide you with the Data Protection registration number.** You will also need to have details about your research project listed on the Divisional Data Protection Holdings or the DCAL holdings. Full details about the information you must provide can be found at: https://www.ucl.ac.uk/pals/intranet/ethics/data-protection

b. Risk Assessment Guidance
Please see the UCL Division of Psychology and Language Sciences Risk Assessment Guidance: https://www.ucl.ac.uk/pals/intranet/ethics/risk-assessment (Password Protected for Divisional staff)

c. Costs
Future costing plans must be made when research data is added to the Research Data Archive. I.e. expansion of server space with UCL Library Services is £800 per 1TB (July 2015).

d. Long-term Digital Curation

i.   All digital material created, owned, or managed by DCAL, will be digitally curated (where data protection laws permit) wherever it is judged to be of sufficiently lasting value under the DCAL Research Data Archive. This will incorporate digital preservation planning to ensure long-term accessibility of the content.

ii.   DCAL will not guarantee to retain all its digitally-curated content indefinitely and unreflectingly (this is also in line with the UCL Retention Schedule). Material which is judged to have outlived its use, and to have little archival value, will be discarded.

iii.   Where data protection laws permit, DCAL will encourage managed approaches to the curation of digital resources by other UCL departments, and will be pleased to advise other UCL departments on the curation of their digital assets.

6. Data Destruction

DCAL follows UCL Information Services' (1) Secure Disposal Guidelines and (2) Information Security Policy, which can be found at the following addresses:

(1) Secure Disposal Guidelines
https://www.ucl.ac.uk/informationsecurity/itsecurity/knowledgebase/securitybaselines/secure_disposal_guidelines
(2) Information Security Policy
https://www.ucl.ac.uk/informationsecurity/policy

At present neither of these documents describe Solid State Hard Drives; these cannot be wiped so need to be destroyed.

Secure shredding and disposal of electronic media can be arranged through Estates; the contact details can be found here: http://wwws-e.ucl.ac.uk/estates/waste/non-hazardous/confidential/

7. Review

This Data Archiving Policy will be revised every two years by the DCAL Data Manager.

Roles for the Data Manager:
- Annually review this Data Archiving Policy and update changes to information, hyperlinks, etc.
- Annually review and amend (as necessary) the DCAL Data Holdings Survey (3.a.ii).
- Annually review and amend (as necessary) the layout or presentation of online content, including the DCAL website and Research Data Archive.
- Review the UCL Research Data Policy every three years to check for amendments or updates – particularly relevant for data re-use and distribution information.
- Every five years review data retention policies.

8. Useful Links

- UCL Data Protection: https://www.ucl.ac.uk/informationsecurity/policy

- UCL Research Data Policy: http://www.ucl.ac.uk/isd/services/research-it/documents/uclresearchdatapolicy.pdf
- UCL Research Registration form: https://www.ucl.ac.uk/finance/legal-insurance/dp-foi
- UCL Retention Schedule: http://www.ucl.ac.uk/library/docs/retention-schedule.pdf
- UCL Research Ethics Committee: https://ethics.grad.ucl.ac.uk/procedures.php
- Division of Psychology and Language Sciences: http://www.ucl.ac.uk/pals/intranet/ethics

9. Contact

DCAL Data Manager
49 Gordon Square
London, WC1H 0PD

Email: dcaldata@ucl.ac.uk

10. Appendices
Appendix 1: Research Consent Form Template

DIVISION OF PSYCHOLOGY AND LANGUAGE SCIENCES ≜ UCL

Informed Consent Form for Participants in Research Studies

This form is to be completed independently by the participant after reading the Information Sheet and having listened to (or seen in BSL) an explanation about the research.

Title of project: XXXXXXXXXX **(corresponding to title of ethics application)**
Specific study: XXXXXXXXXX

This study has been approved by the UCL Research Ethics Committee, Project ID XXXXX
PI: PI name
              PI address
              PI email addresss
              PI telephone number

Experimenter:

NAME EMAIL Tel XXXXXX	

Participant's Statement

I, …………………………………………………………………….., agree that:

- I have read the information sheet and the project has been explained to me orally / in BSL
- I have had the opportunity to ask questions and discuss the study
- I have received satisfactory answers to all my questions or have been advised of an individual to contact for answers to pertinent questions about the research and my rights as a participant and whom to contact in the event of a research-related injury
- Some of my personal details may be passed to UCL for administration purposes
- Research data from this project will be archived securely by UCL
- I understand that I am free to withdraw from the study without penalty if I so wish
- I consent to the processing of my personal information for the purposes of this study only and that it will not be used for any other purpose
- I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998 and UCL Retention Schedule
- **I understand that my responses will be video-recorded and I have completed the separate video consent form**
- Research data from this project may be re-used for other research in accordance with the UCL Research Data Policy and UCL Staff IPR Policy. (Please tick here if you do not want your research data to be used in this way) ☐

Signed: Date:
_____

Investigator's Statement

I, …………………………………………………………….., confirm that I have carefully explained the purpose of the study to the participant and outlined any reasonably foreseeable risks or benefits (where applicable). I have invited this participant to be included in the DCAL Participant Database if they are not already in it.

Signed: Date:
_____

* Remove if no course credit is being offered
** Remove if no video recording is involved

Appendix II: Video Consent Form Template

DIVISION OF PSYCHOLOGY AND LANGUAGE SCIENCES   UCL

Video Consent Form for Participants in Research Studies

This form is to be completed independently by the participant after reading the Information Sheet and having listened to (or seen in BSL) an explanation about the research.

Title of project: <mark>XXXXXXXXXXX</mark> **(corresponding to title of ethics application)**
Specific study: <mark>XXXXXXXXXX</mark>
<mark>Experimenter:</mark>   (insert experimenter name and contact info here).

Your responses will be video-recorded in the course of this study. We will label all data with identifying numbers rather than your name or any other information associated with your identity. However, because face information is important in BSL and in face-to-face communication, it may not be possible to fully conceal your identity. Therefore we will seek your specific consent for the different possible uses of still images or video clips from which you might be recognised. We will only use your still or video images in those circumstances for which you have explicitly given consent.

Please mark "YES" if you give permission for us to use images or brief clips from your video data for a particular purpose, "NO" if you do not give permission.

		Do you give permission for this use?			
1.	**Processing and analysis of your responses by project research staff**	**YES**	☐	**NO**	☐

<u>All of the following uses are strictly optional and will not affect your participation in this study. Please feel free to respond "NO" for any reason. You do not need to provide any explanation to the researcher.</u>

2.	Presentations at academic research conferences	YES	☐	NO	☐
3.	Academic publications reporting the results of these studies, including journal articles, book chapters, technical reports, reports to funding bodies	YES	☐	NO	☐
4.	Educational uses in classroom settings to demonstrate the research methods and/or outcomes	YES	☐	NO	☐
5.	Media reports of the research: a. Print	YES	☐	NO	☐
	b. Television	YES	☐	NO	☐
	c. Internet	YES	☐	NO	☐
6.	Community relations: presentations of the research to groups/organisations within the Deaf community	YES	☐	NO	☐
7.	Presentations where video recording or photography is allowed which could then appear publicly – e.g. the Internet.	YES	☐	NO	☐
8.	Other <mark>[Any "other" must be precisely specified by experimenter here]</mark>	YES	☐	NO	☐

Please note any additional concerns or restrictions on the reverse of this sheet.

Signed: …………………………………………………… Date: ………………………..

Name (Please print) ……………………………………………………………………….

Appendix III: Information Sheet Template

DIVISION OF PSYCHOLOGY AND LANGUAGE SCIENCES   UCL

Information Sheet for Participants in Research Studies

You will be given a copy of this information sheet.

Title of project: XXXXXXXXXX (corresponding to title of ethics application)
Specific study: **XXXXXXXXXX**

This study has been approved by the UCL Research Ethics Committee, Project ID XXXXXX
PI: PI name
                         PI address
                         PI email addresss
                         PI telephone number

Experimenter:

NAME	
EMAIL	
Tel XXXXXX	

We would like to invite you to participate in this research project. You should only participate if you want to; choosing not to take part will not disadvantage you in any way. Before you decide whether you want to take part, please read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or you would like more information.

*In this study, we are interested in your judgements about BSL sentences. Twenty (20) participants will take part in the experiment. All participants are being recruited through the DCAL Participant Database..

In this experiment we are interested to see what aspects of acceptable vs unacceptable sentences in British Sign Language can be detected by people who do not know any BSL. You will see video clips showing BSL sentences. You will be asked to judge if you feel the sentence is "right" or not. Full instructions will be provided orally and in writing.

You will be given some practice sentences at the beginning, and you may ask us any questions before you begin the task. The task should take about 30 minutes to finish. There will be practice sentences and then 4 blocks of sentences, with a break in between each block.

In addition to the BSL Sentence Judgement Task, there is also a screening test for visual-spatial skills. The reason we give the screening test is so that we can report in a research paper that participants all had similar or different visual-spatial skills.*

Your responses will be coded by an identifying code, and not associated to your identity. We may ask you to provide contact details to participate in future related studies, but this contact information will be kept separate from your data.

There are no known risks associated with participation in this experiment. Because there are no known benefits associated with *participation, you will be paid £X as compensation for your time (estimated XX minutes, at a rate of £**£X per hour).)*.

It is up to you to decide whether or not to take part. If you choose not to participate, you won't incur any penalties or lose any benefits to which you might have been entitled. However, if you do decide to take part, you will be given this information sheet to keep and asked to sign a consent form. **Even after agreeing to take part, you can still withdraw at any time and without giving a reason.**

All data will be collected, stored and/ or reused in accordance with the Data Protection Act 1998, UCL Retention Schedule, UCL Research Data Policy, and UCL Staff IPR Policy.

Following your participation in the study, we will provide you with further information regarding the specifics of the study, and you may contact the experimenters if you desire more information [Specific experimenter name and contact information here]; or PI who is supervising these studies, PI email address).

*Replace as appropriate.

** For current DCAL rates, see DCAL staff pages. For student projects see handbook on Moodle for the specific degree programme.

Appendix IV: Data Management Plan Template

The Data Management Plan has been created following guidelines from the Digital Curation Centre, **Please cite as**: DCC. (2013). *Checklist for a Data Management Plan*. v.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans. Available online: http://www.dcc.ac.uk/resources/data-management-plans. This template is recommended in the planning stages, e.g. in preparation of your Data Management Plan for submission with a RCUK research grant proposal.

For the official form, or examples of completed DMPs, please contact the Data Manager.

Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant
Grant Reference Number	Enter grant reference number if applicable [POST-AWARD DMPs ONLY]
Project Name	If this project is funded, state the name exactly as in the grant proposal.
Unique Identifier for Project	E.g. the unique identifier for the BSL Corpus Project is "BSLCP".
Project Description	**Questions to consider:** • What is the nature of your research project? • What research questions are you addressing? • For what purpose are the data being collected or created? **Guidance:** Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.
PI / Researcher	Name of Principal Investigator(s) or main researcher(s) on the project.
PI/ Researcher Contact Details	Email address, address, and telephone number
PI / Researcher ID	E.g. ORCID http://orcid.org/
Project Data Contact	Name (if different to above), telephone and email contact details
Date of First Version	Date the first version of the DMP was completed
Date of Last Update	Date the DMP was last changed
Related Policies	**Questions to consider:** • Are there any existing procedures that you will base your approach on? • Does your department/group have data management guidelines? • Does your institution have a data protection or security policy that you will follow? • Does your institution have a Research Data Management (RDM) policy? • Does your funder have a Research Data Management policy? • Are there any formal standards that you will adopt? **Guidance:** List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.
Data Collection	
This section of the Data Management Plan covers overarching issues concerning data management. Its concerns all qualitative and quantitative data generated by the project. Before completing the section please read the DCAL Data Archiving and Management Plan.	
What data will you collect or create?	**Questions to consider:** • What type, format and volume of data? • Do your chosen formats and software enable sharing and long-term access to the data? • State clearly any access issues. • Are there any existing data that you can reuse? **Guidance:** Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.
How will the data be collected or created?	**Questions to consider:** • What standards or methodologies will you use? • How will you structure and name your folders and files? • How will you handle versioning? • What quality assurance processes will you adopt? **Guidance:** Outline how the data will be collected/created and which community data standards (if any) will be used. Consider how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures. Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.
Documentation and Metadata	

Researchers will be required to provide full metadata together with a description of the datasets which their project generates. The technical arrangements for data management and archiving (including decisions concerning final archiving destination for project datasets; formats for supply of data; licence agreements; IPR etc.) will need to be subsequently agreed with the DCAL Data Manager (dcaldata@ucl.ac.uk). Researchers will be required to meet/liaise with DCAL Data Manager at appropriate intervals to report progress, resolve problems and exchange information.

What documentation and metadata will accompany the data?	**Questions to consider:**	• What information is needed for the data to be to be read and interpreted in the future? I.e. descriptions of data and project data summaries.
		• How will you capture / create this documentation and metadata?
		• What metadata standards will you use and why? (I.e. IMDI).
	Guidance: Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation, and under what conditions it can be accessed. Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards.	

Quality Issues

What procedures will you implement for quality assurance that will be carried out on datasets?	**Questions to consider:**	• Have you documented the calibration of instruments?
		• How will you address the collection of duplicate samples?
		• What are your data entry methods?
		• What data entry validation techniques will you use?
		• What are your methods of transcription?
	Guidance Quality Assurance protocols should be derived for all data collection and processing stages to ensure data are collected as consistently as possible and filenames are standardised.	

Ethics and Legal Compliance

How will you manage any ethical issues?	**Questions to consider:**	• Have you gained consent for data preservation and sharing?
		• How will you protect the identity of participants if required? E.g. via anonymisation.
		• How will sensitive data be handled to ensure it is stored and transferred securely?
	Guidance: Ethical issues affect how you store data, who can see/use it and how long it is kept. Managing ethical concerns may include: anonymisation of data; referral to departmental or institutional ethics committees; and formal consent agreements. You should show that you are aware of any issues and have planned accordingly. If you are carrying out research involving human participants, you must also ensure that consent is requested to allow data to be shared and reused.	
How will you manage copyright and Intellectual Property Rights (IPR) issues?	**Questions to consider:**	• Who owns the data?
		• How will the data be licensed for reuse?
		• Are there any restrictions on the reuse of third-party data?
		• Will data sharing be postponed / restricted, e.g. to publish or seek patents?
	Guidance: State who will own the copyright and IPR of any data that you will collect or create, along with the licence(s) for its use and reuse. For multi-partner projects, IPR ownership may be worth covering in a consortium agreement. Consider any relevant funder, institutional, departmental or group policies on copyright or IPR. Also consider permissions to reuse third-party data and any restrictions needed on data sharing.	

Storage and Backup

How will the data and metadata be stored and backed up during the research?	**Questions to consider:**	• Do you have sufficient storage or will you need to include charges for additional services?
		• How will the data be backed up?
		• Who will be responsible for backup and recovery?
		• How will the data be recovered in the event of an incident?
	Guidance: State how often the data and metadata will be backed up and to which locations. How many copies are being made? Storing data on laptops, computer hard drives or external storage devices alone is very risky. The use of robust, fire-proofed, managed storage provided by university IT teams is preferable. Similarly, it is normally better to use automatic backup services provided by IT Services than rely on manual processes. If you choose to use a third-party service, you should ensure that this does not conflict with any funder, institutional, departmental or group policies, for example in terms of the legal jurisdiction in which data are held or the protection of sensitive data.	
How will you manage access and security?	**Questions to consider:**	• What are the risks to data security and how will these be managed?
		• How will you control access to keep the data secure?
		• How will you ensure that collaborators can access your data securely?
		• If creating or collecting data in the field how will you ensure its safe transfer into your main secured systems?

	Guidance:
	If your data is confidential (e.g. personal data not already in the public domain, confidential information or trade secrets), you should outline any appropriate security measures and note any formal standards that you will comply with e.g. ISO 27001.

Selection and Preservation		
Which data should be retained, shared, and/or preserved?	**Questions to consider:**	• What data must be retained/destroyed for contractual, legal, or regulatory purposes? • How will you decide what other data to keep? • What are the foreseeable research uses for the data? • How long will the data be retained and preserved?
	Guidance: Consider how the data may be reused e.g. to validate your research findings, conduct new studies, or for teaching. Decide which data to keep and for how long. This could be based on any obligations to retain certain data, the potential reuse value, what is economically viable to keep, and any additional effort required to prepare the data for data sharing and preservation. Remember to consider any additional effort required to prepare the data for sharing and preservation, such as changing file formats.	
What is the long-term preservation plan for the dataset?	**Questions to consider:**	• Where e.g. in which repository or archive will the data be held? • What costs if any will your selected data repository or archive charge? • Have you costed in time and effort to prepare the data for sharing / preservation?
	Guidance: Consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving. If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.	

Data Sharing		
How will you share the data?	**Questions to consider:**	• How will potential users find out about your data? • With whom will you share the data, and under what conditions? • Will you share data via a repository, handle requests directly or use another mechanism? • When will you make the data available? • Will you pursue getting a persistent identifier for your data?
	Guidance: Consider where, how, and to whom data with acknowledged long-term value should be made available. The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. If possible, mention earlier examples to show a track record of effective data sharing. Consider how people might acknowledge the reuse of your data.	
Are any restrictions on data sharing required?	**Questions to consider:**	• What action will you take to overcome or minimise restrictions? • For how long do you need exclusive use of the data and why? • Will a data sharing agreement (or equivalent) be required?
	Guidance: Outline any expected difficulties in sharing data with acknowledged long-term value, along with causes and possible measures to overcome these. Restrictions may be due to confidentiality, lack of consent agreements or IPR, for example. Consider whether a nondisclosure agreement would give sufficient protection for confidential data.	

Responsibilities and Resources		
Who will be responsible for data management?	**Questions to consider:**	• Who is responsible for implementing the DMP, and ensuring it is reviewed and revised? • Give details of other researchers/ people who will have access to your data or datasets during and after the project. • Who will be responsible for each data management activity? • How will responsibilities be split across partner sites in collaborative research projects? • Will data ownership and responsibilities for RDM be part of any consortium agreement or contract agreed between partners?
	Guidance: Outline the roles and responsibilities for all activities e.g. data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Consider who will be responsible for ensuring relevant policies will be respected. Individuals should be named where possible.	
What resources will you require to deliver your plan?	**Questions to consider:**	• Is additional specialist expertise (or training for existing staff) required? • Do you require hardware or software which is additional or exceptional to existing institutional provision? • Will charges be applied by data repositories?
	Guidance: Carefully consider any resources needed to deliver the plan, e.g. software, hardware, technical expertise, etc. Where dedicated resources are needed, these should be outlined and justified.	

Appendix V: IMDI Metadata Standards

✓ Denotes mandatory fields

Metadata for DCAL (non-video) data (IMDI standard).

#		Field name	Field description
	Part A (one column for each video file)		
A01	✓	Purpose	Description of why the data was carried out/ project title
A02	✓	Origin	Where the data comes from
A03	✓	Time.References	When the data was created
A04	✓	Geographic.Location	Where the data was compiled (i.e. London, UK)
A05	✓	Creator	List all authors of the data (in citation order)
A06	✓	Access.Conditions	Restriction level (see DCAL restriction levels)
A07	✓	Terms	Any terms of use
A08		Comments	Optional.

Metadata for DCAL video data (IMDI standard).

#		Field name	Field description
	Part A (one column for each video file)		
A01	✓	Purpose	Description of why the data was carried out/ project title
A02	✓	Origin	Where the data comes from
A03	✓	Time.References	When the data was created
A04	✓	Geographic.Location	Where the data was compiled (i.e. London, UK)
A05	✓	Creator	List all authors of the data (in citation order)
A06	✓	Access.Conditions	Restriction level (see DCAL restriction levels)
A07	✓	Terms	Any terms of use
A08		Comments	Optional.

Metadata for DCAL video participant data (IMDI standard).

#		Field name	Field description
	Part A (one column for each video file)		
	Part B (for any Actors not in Sona - e.g. children, atypical adults, or people in stimulus videos)		
B01	✓	Actors.Actor.Code	Recorded on basic metadata sheet
B02	✓	Actor.Age	Recorded on basic metadata sheet
B03	✓	Actor.Sex	M / F
B04	✓	Actor.Education	Highest level of education achieved
B05		Actor.Description	Optional - any other descriptors for actors not mentioned here
B06	✓	Actor.Deafness.Status	Deaf / hard-of-hearing / hearing
B07	✓	Actor.Deafness.Aid Type	None / hearing aid / CI
B08	✓	Actor.Sign Language Experience.Exposure Age	Age of acquisition of SL (years ; months, or birth)
B09	✓	Actor.Sign Language Experience.Acquisition Location	Home from family/ home from tutor/ preschool teachers/ teachers/ family/ beyond home/ friends (more than one may be listed)
B10	✓	Actor.Family.Mother.Deafness	Deaf / hard-of-hearing / hearing
B11	✓	Actor.Family.Mother.Primary Communication Form	Sign / sign supported speech / gesture / mix of signing and speaking / speech only / writing (more than one may be listed)
B12	✓	Actor.Family.Father.Deafness	Deaf / hard-of-hearing / hearing
B13	✓	Actor.Family.Father.Primary Communication Form	Sign / sign supported speech / gesture / mix of signing and speaking / speech only / writing (more than one may be listed)
B14	✓	Actor.Family.Partner.Deafness	Deaf / hard-of-hearing / hearing
B15	✓	Actor.Family.Partner.Primary Communication Form	Sign / sign supported speech / gesture / mix of signing and speaking / speech only / writing (more than one may be listed)
B16		Actor.Education.Age	Age at which school was attended
B17		Actor.Education.School	Name and location of school (if recruited through school)
B18		Actor.Education.School Type	Bilingual home programme / preschool / primary school / secondary school / vocational training / college / university
B19		Actor.Education.Class Kind	Deaf / hard-of-hearing / deaf class in hearing school / individually integrated / hearing only
B20	✓	Actor.Education.Education Model	Bilingual / oral / mixed / sign monolingual / oral with interpreter
B21		Actor.Education.Boarding School	Is the school a boarding school? Y / N

Appendix VI: Profile Details, Part A

Project Details PART A

DCAL is currently collecting data from the last ten years to present to the ESRC in November 2015. As current and former researchers and students that have been involved in various research projects and strands, it would be most helpful if you could provide as full detail as possible relating to the inventory and format of data, by 14 August 2015.

Please complete sections 1, 2, and 3 and return to the Data Archive and Management Officer as soon as possible.

1. Identify ALL (DCAL-funded and –associated) projects you have been involved with:

Example List:
Project 1 Title
Project 2 Title
Etc.

2. Complete this form for <u>each</u> project you have worked on. Prioritise <u>DCAL-Fzunded projects</u>:

Project Title
Give the full title of the project.
Click here to enter text.

Strand/ theme
What strand or theme does the project fall under? Please select one of the following:

☐ Normative Data and Assessment Tools for British Sign Language
☐ Language Development
☐ Face-to-Face Communication
☐ Atypical language
☐ Language Processing
☐ The Deaf Individual and the Community
☐ Sign Language Documentation and Change
☐ Foundations of Communication
☐ Online Measures of Communication
☐ Language and Cognition
☐ Cognitive Control: Executive Functions
☐ Other (please state which): Click here to enter text.

Director/ supervisor
Who is the DCAL Director or Supervisor of this project?
Click here to enter text.

Status
Is this project DCAL-funded or DCAL-associated?

☐ DCAL-Funded   ☐ DCAL-Associated

What is the PI, Funder, and Grant Code if DCAL-Associated?
Click here to enter text.

Size of data
What is the approximate size of ALL (electronic) data associated with this project? (E.g. MB, GB, TB, etc.).
Click here to enter text.

Format of data
Specify ALL the kinds of data related to this project, i.e. paper, external hard drive, DVDs, scans, ELAN files, etc. Include research data as well as stimuli, metadata, etc.
Click here to enter text.

Video data
Are any of your data video data that cannot be anonymised?

☐ Yes   ☐ No

Location of data
Describe clearly where all the data for this project is stored. Identify multiple locations if necessary.
Click here to enter text.

Consent and Ethics
We need to know what consent you acquired for this project. Attach copies of all (blank) consent forms used for the project to this document.

Anonymisation of non-video data

Has non-video data been anonymised?   ☐ Yes   ☐ No

Appendix VII: Profile Details, Part B

Project Details PART B

Ensure that you have completed sections 1, 2, and 3 of the Project Details PART A form, and have returned this to the Data Archive and Management Officer.

Before continuing, read the 'Guidelines to Support the Preparation of Data for ESRC Submission' instructions (see page 6, below).

PART B: Sections 1 (the Project Outline) and 2 (Attachments) are to be completed and submitted when you deliver all of your project data; work with the 'Guidelines to Support the Preparation of Data for ESRC Submission' guide. This can be submitted at a later date, but ensure that the Data Archive and Management Officer knows approximately when to expect your submission.

1. Project Outline:

Project Title
Give the full title of the project.
Click here to enter text.

Project Code
Specify a single project title to be used with data in the archive; a unique acronym since some will be long. I.e. British Sign Language Corpus Project = 'BSLCP'.
Click here to enter text.

Key references/ papers
Identify one or two key papers which have been published through this project if relevant.
Click here to enter text.

Identify all author(s) of the data (in order)
This will be used for the citation of the data.
Click here to enter text.

Project Description:
Short summary (150 characters) about the project that these data were used for.
Click here to enter text.

Abstract:
Paragraph describing the project that these data were used for.
Click here to enter text.

Summary of all archived data:
Include wherever possible, collected data, anonymised metadata about participants, stimuli, project information sheet, and blank consent form(s).
- *You may find it useful to group your data into certain folders and describe the data sets in each folder.*
- *If you need to explain any labels or codes you have used (and who should be contacted to decode that information), this is a good place to do so.*

Click here to enter text.

Access conditions/ consent
Relevant wording from info sheet or consent form about sharing/ storing/ destruction of data: For example: "Your responses and test scores will be stored on the DCAL server and only our research staff will have access."

Click here to enter text.

2. Attachments to data and/ or datasets

Permission/ restriction level
Specify the suggested level of permission for your data, with justification. Please maximise sharability, which may involve setting different levels of permission for different subtypes of your data. Key for permission levels:

1   Open Access – available publically to all users. Assign to stimuli. (Specify any conditions or terms of use, e.g. CC license, etc. CC BY-NC-SA is recommended.)

2   Restricted to researchers – only researchers who have been granted access (from the project's PI by contacting DCAL) will have access to the data.

3   Restricted to DCAL – only DCAL-funded researchers (who have been granted access from the project's PI) will have access to the data.

4   Locked. Only the PI(s) will have access to the project data. Assign to Child or Patient data.

Data Label	Restriction Level	Justification (for Restriction Level)
File name #	2	Consent permits sharing only for further research.
Folder name #	1 - CC BY-NC-SA	Stimuli – can be shared and reused.
Folder name #	4	Child video data.

Metadata

The following metadata is required by the ESRC to accompany each item of data or dataset (a set of data that holds similar and 'grouped' properties, which may be a folder or subfolder). Please attach this metadata to all of your data (MS Excel/ .xls format is acceptable).

Example:

Purpose	Description of why the data was carried out/ project title	BSL Grammaticality Judgement Task
Origin	Where the data comes from	DCAL (London, UK)
Time references	When the data was created	Oct 2012
Geographic location	Where the data was compiled (i.e. London, UK)	London, UK
Creator	List all authors of the data (in order)	Cormier, K., Schembri, A., Vinson, D. Orfanidou, E.
Access Conditions	Restriction level (see point below)	1
Terms of use of data	Any terms of use	CC BY-NC-SA or © UCL DCAL
Comments (optional)	Optional field	-

- If you have a folder full of information where the metadata will be the same across all files and subfolders, it is okay to include one .xls file which is attributed to all of those.
- If, however, there are files where the metadata information varies, you need to ensure that those files have appropriate and relevant .xls files attached. If you have one folder where generally all the metadata will hold the same information besides a few exceptions, you can just create one .xls document and then comment on those exceptions in the 'Comments' field of the metadata form.
- If there are just too many files for you to logistically attach metadata to (i.e. if you're in a position where you would need to create a .xls document for each file), it might make sense to give more general metadata information such as "September 2013" instead of specific dates like "12 September 2013", etc.
- Just have a think about what: (1) makes logistical sense, (2) will be important for future researchers, and (3) you can clearly organise and explain.

Guidelines to Support the Preparation of Data for ESRC Submission

Once you have collated all of your research data associated with your projects, you must prepare and organise it for submission to the ESRC. This must be completed by November 2015 to meet ESRC requirements; <u>if you do not complete this on time, the ability of DCAL researchers to get further funding could be affected</u>.

1. Ensure that you have completed sections 1, 2, and 3, of the Project Details PART A form, and have returned this to the Data Archive and Management Officer.
2. Organise your data in folders and subfolders (datasets) according to each project. It is largely up to you how to organise the data, but you must ensure that it is easy to navigate, with separate folders for stimuli and research data.
3. Data must have clear and basic labels which other researchers can easily understand; all files (in any format), folders, and sub-folders must have understandable and simple labels. All labelling must be consistent. Attach a 'Read Me' document if it is necessary to describe data e.g. in terms of how it was collected, coded, analysed, etc.
4. Sections 1 and 2 of the Project Details PART B form must be completed fully, and submitted, alongside all research data.

Additional Information:
- Example of projects ready for submission can be found on the shared DCAL drive:
  Shared (P:\) > BSLGJT for archiving and *Shared (P:\) > BSLSRT for archiving*.
- All hard-copy data will be destroyed so you *must* ensure that relevant research data is digitised. Contact the DCAL Data Archive and Management Officer if you need help with this.

Checklist
☐ All project data (organised as set out above)
☐ Data has been anonymised where possible
☐ Part A and B of the Project Detail form

If you have any further questions or information relating to your data or this form, please let me know. Once completed, return to: datamanager@ucl.ac.uk. Thank you.