



**UK Statistics
Authority**

**Fourth Meeting of the
National Statistician's Data Ethics Advisory Committee**

Minute, Agenda and Papers

Tuesday 19 April 2016

11:00 – 15:40

Room 1103-1105, UK Statistics Authority
Titchfield

National Statistician's Data Ethics Advisory Committee

Minute

Tuesday, 19 April 2016
1103-1105, Titchfield, Fareham

Present

Members

Mr Ian Cope (Chair)
Mr Robert Bumpstead (Deputy Chair)
Ms Annie Hitchman
Dr Dean Machin
Ms Isabel Nisbet
Mr Osama Rahman
Professor Martin Severs
Mr Hetan Shah

UK Statistics Authority

Mr Adil Deedat
Dr Simon Whitworth

Office for National Statistics

Ms Carolyn Watson (for item 3)
Ms Lucy Vickers (for item 3)
Mr Andy Teague (for item 4)
Mr Steve Woodland (for item 5)
Ms Jen Woolford (for item 5)
Mr Pete Brodie (for item 6)
Ms Emma Timm (for item 6)
Mr Neil Bannister (for item 9)

Welsh Government

Mr Stephen Hughes (for Item 7)

Scottish Government

Mr Albert King (for item 8)
Mr Gerald Donnelly (for item 8)

Apologies:

Mr Colin Godbold
Mr Neil McIvor
Ms Marion Oswald

1. Minutes and matters arising from the previous meeting

- 1.1 The Chair welcomed the National Statistician's Data Ethics Advisory Committee (NSDEC) to the Authority's site in Titchfield and to the committee's fourth meeting.
- 1.2 The Chair informed the meeting that the minute of the third meeting had been agreed and signed off by correspondence. The minute, agenda and papers from the last meeting are now published on the [UK Statistics Authority website](#).

- 1.3 Progress with actions from the previous meetings was reviewed. Most actions were complete or in progress and the secretariat updated the committee accordingly.
- 1.4 The Chair updated the committee on progress with projects which had been deliberated on at previous meetings.
- 1.5 The meeting heard that measures were still being considered as to how the ethnicity from names tool could be made accessible to those who have a clear public benefit from using the tool under strictly controlled arrangements. It is likely that the mode to accessing the tool would be through the ONS Approved Researcher route. Digital Services, Technology and Methodology Directorate in ONS are working with Information Management (IM) business partners to consider ways in which the tool could be disseminated without risk of misuse. The Chair informed the committee that if a technical solution could not be found, then the proposal would not go ahead.
- 1.6 The proposal for the use of deaths registration data as a sampling frame for the VOICES survey received minor revisions at the last meeting. NSDEC were informed that in light of recommendations, privacy notices had been provided to the General Registrars Office who was in turn circulating these to local authority registry offices in England and Wales. This will inform those registering a death or other vital event that their data may be used for statistical purposes. Other recommended revisions included considering ways in which those who had previously been included in the sample of the VOICES survey and had declined to participate can be removed from any future sampling frame. The Life Events and Population Sources Division in ONS are working with IM partners to explore how such a system could be developed. The final recommendation was for the information leaflet to be restructured with more information provided about ONS and the purpose of the survey. Implementation of this recommendation will be taken forward following NHS England's options appraisal for collecting data on quality of end of life care.
- 1.7 Further to discussion on the Twitter project at the last meeting, the matter was referred to the National Statistician who suggested that the Twitter project run its course until the funding ended (March 2016). Thereafter, any proposal to acquire more Twitter data and utilise new methodologies, will require a formal application to NSDEC.

2. Chair's report

- 2.1 The Chair informed the meeting that Charles Bean's review of economic statistics had reported. The report was circulated to members at the time of publication. The Chair made reference to a number of relevant recommendations, including stronger powers for accessing data and the need for an independent person or body to provide ethical oversight.
- 2.2 The meeting heard that the public dialogue on the ethics of data science in Government that is being co-ordinated by Government Digital Service (GDS) and Cabinet Office and part funded by ONS, will shortly be releasing its findings. The Secretariat is liaising with relevant colleagues within these departments and will arrange for them to come and present the findings of the research at a future meeting.
- 2.3 The Chair invited Mr Shah to inform NSDEC of his role in giving evidence to the Commons Science and Technology select Committee on the "Big data dilemma" in December 2015. Mr Shah informed NSDEC that his suggestion of a council for data ethics would be for a more overarching role than that of NSDEC, with the aim of considering broad data ethics issues across government over the next 15 to 20 years.
- 2.4 The Secretariat updated the committee on the NSDEC application form and guidance which currently appears as a downloadable document on the NSDEC web pages. As a means to better facilitating researcher applications, the application form and guidance

will be embedded in to a portal and will provide the guidance alongside questions. Members welcomed the change but suggested that a free standing application form is still available so that prospective applicants are able to see, in advance of completing an application, exactly what information will be required.

- 2.5 The Chair informed the meeting that he will be presenting at the forthcoming “Trust, Risk, Information and the Law” conference at the University of Winchester. The presentation will cover the use of big data techniques and new data sources in the production of national statistics and the opportunities and the ethical and governance challenges associated with these. It was suggested that a generic set of slides about NSDEC be circulated to members to use for when they talk about the work of the committee. Members were asked to keep the Secretariat informed of when they speak about the work of the committee.
- 2.6 The Chair also informed NSDEC of a paper he presented to the National Statistician’s Inter Administration Committee on the role and work to date of NSDEC. The Chief Statisticians in the Devolved Administration’s welcomed NSDEC’s role and were keen to engage with it.

Action: Secretariat to ensure free standing application form is remains available to applicants to view on creating an online application process.

Action: Secretariat to develop generic slides for members to use when speaking at conferences.

3. Data Sharing

- 3.1 The Chair invited Ms Vickers and Ms Watson from the ONS Administrative Data Division to discuss with NSDEC new data shares being sought for research and statistical purposes and to inform members about the measures in place to protect administrative data held by ONS.
- 3.2 The meeting heard that the secure research environment in which administrative data is linked has received accreditation by CESG, the information assurance branch of Government Communication Head Quarters (GCHQ).
- 3.3 Ms Vickers then provided NSDEC with an update on new data shares at **Annex A**. Members suggested that more information needed to be provided on data shares. Along with the name of the data set and the department from which the data is being requested, NSDEC asked for the following to be included in future updates on shares:
 - i. what data is being sought, for example whether demographic, economic etc;
 - ii. the level of the data that is being sought, for example whether aggregate or record level and whether the data is identified or de-identified;
 - iii. the legal gateway by which the data has been or will be acquired; and
 - iv. the public benefit or public good in acquiring and using that data.
- 3.4 It was suggested that the business cases developed to acquire new data sources be communicated to NSDEC at future meetings.

Action: Secretariat to work with the Administrative Data Division to ensure that future data share updates include information requested by the committee.

4. **Visit to the Census Transformation Programme**

- 4.1 The Committee were given a tour of the Census smart working area. Mr Teague from the Census Transformation Programme team provided NSDEC with an overview of ONS's ambition for an administrative data Census, which in the future will see greater use of government and potentially other data sources. ONS will be publishing an annual assessment of ONS's ability to move to an administrative data census after 2021.

5. **Draft Safeguarding policy [NSDEC(16)04]**

- 5.1 Mr Woodland from ONS Social Survey Division and Ms Woolford from ONS Life Events and Population Sources Division were welcomed to the meeting. Mr Woodland introduced the paper and set out the issues.
- 5.2 The meeting heard that when ONS is commissioned to undertake surveys, commissioners are asking ONS for a safeguarding policy.
- 5.3 NSDEC considered the proposals outlined in the draft safeguarding policy paper, which they noted was still a work in progress. It was suggested that greater clarity was required around legal constraints, and that it was important for NSDEC to focus on the ethical aspects of the work.
- 5.4 It was also suggested that where ONS is the data controller then a safeguarding policy for ONS is appropriate. Where ONS is a data processor, ONS should be adhering to the safeguarding policy of the data controller.
- 5.5 The use of real life case studies was seen to be a useful way of defining a safeguarding policy. As new instances of safeguarding issues present themselves, case studies could be updated accordingly. Examples of good practice could also be drawn from lone worker policies used by the emergency services.
- 5.6 Consideration was given to the interest of particular groups including children and vulnerable adults should be put before anything else.
- 5.7 There was a need for the safeguarding policy to address the different modes (e.g. face to face interview or postal survey response) by which safeguarding issues arise.
- 5.8 The Committee agreed that the draft safeguarding policy should be developed further before further consideration at the next meeting. The discussions from the meeting will also be shared with the Government Statistical Service Statistical Policy and Standards Committee.

Action: The Deputy Chair in his capacity as Chief of Staff to the National Statistician to work with the National Statistician in supporting Ms Woolford and Mr Woodland in further developing the draft safeguarding policy.

6. **Application: Use of paradata to select participants for follow up research [NSDEC(16)05]**

- 6.1 Mr Brodie from Survey Methodology and Statistical Computing Division in ONS introduced the item and reaffirmed the definition of paradata. Mr Brodie informed the Committee that collecting paradata from the Monthly Wages Salaries Survey facilitated better questionnaire design.
- 6.2 NSDEC suggested that the project could proceed subject to sight of the consent statement, which is currently being developed. The consent statement should include all proposed uses of the paradata and cite appropriate security measures.

Action: Mr Brodie to provide the consent statement for circulation to members for consideration along with the information sheet referenced within the application, before the next meeting.

7. Application: Using Welsh Schools data to understand the effects of bilingualism on academic performance [NSDEC(16)06]

- 7.1 The Chair invited Mr Hughes to provide an overview of this item. Mr Hughes informed the Committee that the proposal is not a joint project between Welsh Government and the University of Bath, as such, but rather the Welsh Government's role is as a data controller and data provider.
- 7.2 Mr Hughes explained that schools data had previously been shared with other institutions, however this would be the first time information relating to names would be shared. The meeting heard that the names would not be identifiable and would be aggregated into groups informing whether the name was Welsh, not Welsh or had an unidentified origin.
- 7.3 Members agreed that due to the aggregate and de-identified nature of this data, there was no issue sharing the data with the University of Bath. Members were unsure, however, whether the research proposal could inform accurately of a causal link between bilingualism and academic performance, but agreed that this was beyond scope.
- 7.4 Members also sought clarity on the fair processing notice and whether this covered the use of data for the purposes outlined. It was suggested standard Government Statistical Service processes should be followed and that processing should conform to Chapter 7 of the Information Commissioner's Office.
- 7.5 NSDEC agreed to approve the proposal for Welsh Government to share data with the University of Bath subject to ethical approval being granted by the University of Bath's ethics committee.

Action: Secretariat to follow up with Welsh Government and to circulate relevant ethical approval documents from the University of Bath's ethics committee

8. Application: Scottish Government, Education identifiers-read through [NSDEC(16)07]

- 8.1 The Chair invited Mr King and Mr Donnelly to present this item. The meeting heard that the proposal would streamline the linkage of Scottish education data to Scotland's spine of records, in a way that protects and strengthens privacy.
- 8.2 NSDEC agreed that the proposed method was a sensible approach to ever increasing demands for linked education data.
- 8.3 There was unanimous support for the proposal, and members suggested that the methods be shared with other relevant parties including the Administrative Data Research Network and the Farr institute.
- 8.4 The Committee advised the National Statistician to approve the research.

Action: Secretariat to follow up with Scottish Government to ensure that the methods employed are disseminated to relevant parties.

9. Paediatric national survey of bereaved people [NSDEC(16)08]

- 9.1 Mr Bannister from Life Events and Population Sources Division introduced the item and explained that the proposal was similar to the National Survey of Bereaved people, which was reviewed at the last meeting, but related to end of life care for children. ONS's involvement in the proposal would be to administer a survey designed by the University of Southampton using the deaths register as a sampling frame.
- 9.2 The meeting heard that the proposal was at an early stage of development and that advice from the Committee at this early stage would be useful in refining the proposal.
- 9.3 NSDEC noted that use of deaths data had been approved for sampling people who had registered a death to understand end of life care for adults. However there were greater sensitivities when asking those who have registered a child's death (most likely a parent) about the end of life care for the child.
- 9.4 Members were supportive of the proposal, however in order for the NSDEC to reach a fully informed decision, the information leaflet and the letter, which would be sent to registrants, would need to be reviewed by the Committee. It was also suggested that the current number of questions proposed be reviewed to ensure burden was minimised.
- 9.5 Mr Bannister informed the Committee that drafts of the information leaflet and letter were in production. In addition, he informed members that the number of questions was currently being reviewed and may be reduced.
- 9.6 The section of the questionnaire relating to free text data was also discussed along with its proposed use by the University of Southampton in undertaking qualitative research. Members agreed that more information was needed to understand what this qualitative research would entail and how its insight could be beneficial to the public.
- 9.7 It was suggested that, one reminder, following failure to respond would be more appropriate than two reminders. In light of earlier comments on the draft safeguarding paper it was suggested that NSDEC has sight of the University of Southampton's safeguarding policy.
- 9.8 Members suggested that the proposal be considered at a future meeting with the following points addressed so that Committee can provide fully informed advice:
- i. the information leaflet and letter to be sent to sampled registrants to be included with the application;
 - ii. consider reducing the number of questions in the survey;
 - iii. consider whether one reminder should be sent to registrants;
 - iv. sight of the University of Southampton's safeguarding policy; and
 - v. sight of the University of Southampton's ethics committee's deliberations on the research proposal with specific relation to the aspects of the proposal for which the University of Southampton are responsible.

Action: Mr Bannister to address committee comments and return to a future meeting with an amended proposal reflecting member's comments and recommendations.

10. Any other business

- 10.1** Members discussed future agenda items. It was suggested that the Code of Practice for Official Statistics, which is currently undergoing review, could be considered at a future meeting. Other future agenda items could include the ethics of hack-a-thon. As NSDEC completes its fourth meeting and final meeting of its first year of operation, it was agreed that at the next meeting the Committee undergoes a self assessment.

Action: Secretariat to develop a self assessment for the next meeting and to consider other items suggested for inclusion at future meetings.

UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE

Agenda

**Tuesday 19 April
Room 1103-1105, Titchfield
11:00am – 3:40pm**

Chair: Mr Ian Cope
Apologies: Ms Marion Oswald
Mr Neil McIvor
Mr Colin Godbold

(11:00am to 12:00pm)

1 11:00am	Minute and matters arising from the previous meeting	Mr Ian Cope
2 11:05am	Chair's report	Oral Report Mr Ian Cope
3 11:20am	Data shares Annex A: acquisition of new data sources	Oral Report Ms Lucy Vickers Ms Carolyn Watson

Lunch (12:00pm to 12:30pm)

(12:30pm to 3:40pm)

4 12:30pm	Visit to Census Transformation Programme	Mr Andy Teague
5 1:00pm	Draft safeguarding policy	NSDEC(16)04 Ms Jen Woolford Mr Steve Woodland
6 1:30pm	Using paradata to select participants for follow up research	NSDEC(16)05 Mr Pete Brodie Ms Emma Timm
7 2:00pm	Using Welsh schools data to understand the effects of bilingualism on academic performance	NSDEC(16)06 Mr Stephen Hughes
8 2:30pm	Scottish Government - education identifiers read through	NSDEC(16)07 Mr Albert King Mr Gerald Donnelly
9 3:00pm	The paediatric national survey of bereaved people	NSDEC(16)08 Mr Neil Bannister
10 3:30pm	AOB	

Next meeting: Tuesday 5 July 2016

Location: One Drummond Gate London

National Statistician's Data Ethics Advisory Committee**Minute**

**Wednesday, 27 January 2016
Boardroom, Drummond Gate, London**

Present**Members**

Mr Robert Bumpstead (Deputy Chair)
Mr Colin Godbold
Ms Annie Hitchman
Dr Dean Machin
Mr Neil McIvor
Ms Isabel Nisbet
Ms Marion Oswald
Mr Osama Rahman
Mr Hetan Shah

UK Statistics Authority

Mr Adil Deedat
Dr Simon Whitworth
Ms Natalie Shorten
Mr Ross Young (for item 3)

Office for National Statistics

Mr Owen Abbot (for item 4)
Mr Alistair Calder (for item 5)
Mr Nick Stripe (for item 6)
Mr Mark Gautrey (for item 6)
Ms Helen Colvin (for item 6)
Ms Jane Naylor (for item 7)
Mr Thomas Smith (for Item 7)

Apologies:

Mr Ian Cope (Chair)
Professor Martin Severs

1. Minutes and matters arising from the previous meeting

- 1.1 The minute from the last meeting was agreed by correspondence, signed off by the Chair and is now published on the [UK Statistics Authority Website](#).

Chair's report

Mr Ian Cope

Data shares

Oral report

Ms Lucy Vickers & Ms Carolyn Watson

List of Annexes

Annex A Acquisition of new data sources, Carolyn Watson, 7 April 2016

Annex A- Acquisition of new data sources

Data Set	Information
PAYE and Benefits information	Update: Following the delivery of data for feasibility research ONS has started discussions around an on-going supply of data which could be used more widely across the office
All England Education Dataset	Update: ONS have received data which will allow them to carry out research for BIS under s22 of the SRSA.
Health Demographic	User requirements being collated and refined.
Driver and Vehicle Licensing Agency data	Aggregate data have been provided detailing the number of people who have updated their details electronically by postcode. No further acquisition activity is currently planned.
TV Licensing	Work starting to understand the legal position with regards to the data. Could provide information for an address register.
Valuation Office Agency Data	Data due to be delivered early in the next financial year to support Census Transformation Programme (CTP) and Social Survey Division (SSD) access to VOA data.
National Border Target Centre	Data is to be used to support feasibility work in improving the International Passenger Survey (IPS) sampling and weighting frame. A series of unrounded aggregate data extract reports on inbound and outbound data between the dates of 1 October to 7 October 2015 was delivered on 19 November 2015. Further data extracts may be requested depending on the results of the initial feasibility work.
Corporation Tax	Access is being sought to data used to support the calculation of corporation tax liabilities. This information will support a number of business areas and, subject to successful feasibility work, will lead to improvements in GDP, the sampling of businesses for surveys and the quality assurance of existing outputs.
Data from the Financial Conduct Authority	Access to these data are being sought to support the development of new Flow of Funds outputs based upon how revenue moves between different business entities.

Visit to the Census Transformation Programme
Mr Andy Teague

UK Statistics Authority
National Statistician's Data Ethics Advisory Committee
NSDEC(16)04
Draft safeguarding policy

This paper is in draft and will be published shortly

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(16)05

Using paradata to select participants for follow-up research

Purpose

1. The Data Collection Methodology (DCM) team in the Office for National Statistics (ONS) is exploring the use of paradata as a means of further understanding and improving user experiences when responding to online surveys. As ONS moves to greater electronic data collection, DCM would like to use paradata as a way of selecting respondents for follow up to understand their online experience when responding to ONS surveys.
2. A specific application for using paradata from the Monthly Wages and Salaries Survey (MWSS) is presented at **Annex A**.

Recommendations

3. Members of NSDEC are invited to:
 - i. consider the case for using paradata to select online survey respondents for follow up to better understand respondent experiences when completing ONS questionnaires;
 - ii. review the application presented at **Annex A** and advise the National Statistician to:
 - a) approve the proposal and allow it to proceed;
 - b) approve the proposal subject to minor revisions;
 - c) recommend major revisions to the proposal and request that it be resubmitted to a future meeting once implemented; or
 - d) reject the proposal advising that the research be stopped from proceeding; and
 - iii. in light of advice at **Annex A**, offer guidance on whether the approach outlined in this application can enable the same research to be undertaken on other ONS business and social surveys and the census.

Background

4. ONS is undertaking a programme of work to move business and social surveys, and the census, to electronic modes of data collection. This focus on electronic data collection has presented opportunities to access additional data; namely paradata.
5. Paradata are automatic data collected about the survey data collection process captured during computer assisted data collection and include call records, interviewer observations, time stamps, keystroke data, travel and expense information, and other data captured during the process.
6. Analysis of paradata supports continuous improvement of questionnaires, associated systems and the wider services that support them. The use of paradata and follow up of respondents to whom the paradata relate could reduce the time and cost burden on those responding.

Pete Brodie and Emma Timm, Methodology Division, ONS, 4 April 2016

List of Annexes

Annex A Application: Using paradata to select participants from the Monthly Wages and Salaries Survey for follow-up research, ONS Methodology, Pete Brodie & Emma Timm, 4 April 2016



National Statistician's Data Ethics Advisory Committee

Application for Ethical Review

6.1

The Application Process

This is an application form for applying for ethical review from the National Statistician's Data Ethics Advisory Committee (NSDEC). You should use the additional guidance when completing this form.

The application form should be completed in **plain English** which is understandable to lay members and all abbreviations should be explained the first time they are used. The form should contain sufficient information to ensure a thorough ethical review can take place.

Please word process the form using Arial or Times New Roman font, size 11. Where necessary expand text boxes on the form to accommodate answers, but ensure word counts are adhered to where specified.

Where sections are not relevant to your study please mark as N/A.

On completion the responsible owner should sign the application form and send to:
nsdec@statistics.gsi.gov.uk



Section A
Application Details

A1	Responsible Owner		
Full Name: [REDACTED]		Position: [REDACTED] Methodology	
Address: [REDACTED] Office for National Statistics, [REDACTED] [REDACTED] [REDACTED]		Email: [REDACTED]	
		Telephone: [REDACTED]	
		Organisation: Office for National Statistics	
<p>Declaration to be signed by the responsible owner</p> <p>I have met with and advised the applicant on the ethical aspects of this project design <i>(applicable only if the responsible owner is not the Applicant)</i>.</p> <p>I understand that it is a requirement for all researchers accessing the data to have undergone relevant training and to have either relevant security clearances or approved researcher status in order to access the data.</p> <p>I am satisfied that the research complies with current professional, departmental and other relevant guidelines.</p> <p>I will ensure that changes in approved research protocols are reported promptly and are not initiated without approval by the National Statistician's Data Ethics Advisory Committee.</p> <p>I will provide notification when the study is complete if it or fails to start or is abandoned.</p> <p>I will ensure that all adverse or unforeseen problems arising from the research are reported in a timely fashion to the National Statistician's Data Ethics Advisory Committee.</p> <p>I will consider all advice received from the National Statistician's Data Ethics Advisory Committee and should I be unable to implement any of the recommendations made, I will provide reasoning in writing to the Committee.</p>			
<p>Print Name: [REDACTED]</p> <p>Signature: [REDACTED]</p> <p>Date: 4th April 2016</p>			



A2 Applicant Details (if applicant is not the responsible owner)	
Full Name: Emma Timm	Position [REDACTED] Data Collection Methodology
Address: [REDACTED] Office for National Statistics, [REDACTED] [REDACTED] [REDACTED]	Email: [REDACTED]
	Telephone: [REDACTED]
	Organisation: Office for National Statistics (ONS)

A3 Project Information	
Project Title: Using paradata to select participants for follow-up research	
Start Date: 26/04/2016	End Date: Continuous
Project Sponsor (select all that apply)	
<input checked="" type="checkbox"/> ONS <input type="checkbox"/> ADRN <input type="checkbox"/> GSS <input type="checkbox"/> Collaboration <input type="checkbox"/> Other (Please specify).....	

A4 Collaboration and Sponsors	
List of Collaborators/Sponsors	Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)
None	



A5	Proposed Site of Research (select all that apply)
<p>Where will the research take place?</p> <p><input checked="" type="checkbox"/> ONS</p> <p><input type="checkbox"/> VML</p> <p><input type="checkbox"/> HMRC Data Lab</p> <p><input type="checkbox"/> ADRC-England</p> <p><input type="checkbox"/> ADRC-Northern Ireland</p> <p><input type="checkbox"/> ADRC-Scotland</p> <p><input type="checkbox"/> ADRC-Wales</p> <p><input type="checkbox"/> Other (please specify).....</p>	
<p>Is this a secure site?</p> <p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p>	

Section B

Project Details

B1	Please provide a brief high level summary of the research giving necessary background (<i>max 500 words</i>)
<p>Office for National Statistics (ONS) is undertaking a programme of work to move business and social surveys, and the census to electronic modes of data collection. This focus on electronic data collection has presented opportunities to access additional data; namely paradata.</p> <p>“Paradata are automatic data collected about the survey data collection process captured during computer assisted data collection and include call records, interviewer observations, time stamps, keystroke data, travel and expense information, and other data captured during the process” (Kreuter, Couper & Lyberg, 2010¹).</p>	

¹ Kreuter, F., Couper, M., & Lyberg, L. (2010). The use of paradata to monitor and manage survey data collection. *Section on Survey Research Methods – JSM*, 83-296.

Analysis of these paradata supports continuous improvement of questionnaires and the systems and wider service that support them. During an online pilot of the Monthly Wages and Salaries Survey (MWSS), the analysis of paradata assisted in the identification of issues with the questionnaire and associated processes for completion. In one interesting case, a respondent showed a pattern of moving through the survey, reaching the last question, and then, at a later date, going back through all of the questions from the beginning, before submitting.

By matching these paradata to ONS call centre records; it was found that the respondent called to request a specimen questionnaire. They explained that to complete the questions within the survey, they required data from different HR teams within their business. The only way of knowing what was contained within the survey was for the respondent to enter false data into the questionnaire to view all questions. When actual data was available, the respondent completed and submitted the survey. This insight demonstrated how ONS could significantly reduce burden on respondents by providing an option to preview the questionnaire. Consequently, a research project is being scoped to fully understand what information respondents need prior to accessing the survey, and to test different solutions.

Additional call information is not always available. The aim of the proposed research is to identify cases of interest through analysing the paradata of respondents who have completed/attempted to complete the MWSS online. Respondents' paradata will be linked, where possible, to other data held on ONS systems (e.g. call records² and Interdepartmental Business Register³ (IDBR)). Subject to businesses not being in enforcement for non-completion of any statutory ONS surveys, the research team(s) will attempt to recruit them to take part in follow-up research. The methods used in this follow-up research will predominantly involve semi-structured interviews in person or by telephone to examine:

- i. the reasons behind their interesting paradata
- ii. their general experience of registering for and completing the survey online
- iii. respondent needs for specific aspects of the system or survey

The data gathered through the follow-up research will be used to:

- i. make improvements to questionnaires
- ii. make improvements to the overall service (from registration through to submission)
- iii. inform topics for further in-depth research

² **Call records:** MWSS respondents may have called ONS' Business Data Division (BDD), to ask for assistance or provide feedback while registering for the service and completing the survey. Equally, call records may relate to calls that BDD has made to MWSS respondents to chase responses and/or query data. These data provide additional context for the respondent's experience while using the RRM / eQ.

³ Data from IDBR used to supplement paradata include: registered employment; Standard Industrial Classification; selection type and history; etc.

B2	Data Use																																					
	<table border="1"> <thead> <tr> <th rowspan="2">Type of data</th> <th colspan="4">Data Level <i>Please specify the name of the data set</i></th> </tr> <tr> <th>Aggregate Data</th> <th>Identifiable Data</th> <th>De-identified personal data</th> <th>Anonymised/ pseudo anonymised</th> </tr> </thead> <tbody> <tr> <td>Administrative data <i>(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)</i></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Big Data <i>(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)</i></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Survey Data <i>(please specify e.g. LFS, BRES, etc in the relevant options adjacent)</i></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Census Data <i>(please specify year, e.g. Census 2011 in the relevant options adjacent)</i></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Other <i>(please specify e.g. Ordnance Survey Address register in the relevant options adjacent)</i></td> <td></td> <td>ONS records of incoming and outgoing calls from survey respondents.</td> <td></td> <td>Paradata from products used to maintain the electronic data collection (EDC) service⁴.</td> </tr> </tbody> </table>				Type of data	Data Level <i>Please specify the name of the data set</i>				Aggregate Data	Identifiable Data	De-identified personal data	Anonymised/ pseudo anonymised	Administrative data <i>(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)</i>					Big Data <i>(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)</i>					Survey Data <i>(please specify e.g. LFS, BRES, etc in the relevant options adjacent)</i>					Census Data <i>(please specify year, e.g. Census 2011 in the relevant options adjacent)</i>					Other <i>(please specify e.g. Ordnance Survey Address register in the relevant options adjacent)</i>		ONS records of incoming and outgoing calls from survey respondents.		Paradata from products used to maintain the electronic data collection (EDC) service ⁴ .
Type of data	Data Level <i>Please specify the name of the data set</i>																																					
	Aggregate Data	Identifiable Data	De-identified personal data	Anonymised/ pseudo anonymised																																		
Administrative data <i>(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)</i>																																						
Big Data <i>(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)</i>																																						
Survey Data <i>(please specify e.g. LFS, BRES, etc in the relevant options adjacent)</i>																																						
Census Data <i>(please specify year, e.g. Census 2011 in the relevant options adjacent)</i>																																						
Other <i>(please specify e.g. Ordnance Survey Address register in the relevant options adjacent)</i>		ONS records of incoming and outgoing calls from survey respondents.		Paradata from products used to maintain the electronic data collection (EDC) service ⁴ .																																		

B3	How will information be kept confidential and data kept secure? <i>(max 500 words)</i>
<p>Paradata are collected during the process of completing electronic questionnaires. These data are collected and stored securely by the 'products' that are used to maintain the electronic data collection service. For the MWSS pilot the product used to collect paradata was the Response and Respondent Management System (RRM). In the MWSS pilot the RRM product enabled respondents to register to the service; complete the MWSS survey; and enabled survey and paradata to be stored for later extraction to other systems and programmes for analysis.</p>	

⁴ Examples of EDC products: Response and Respondent Management system (RRM); electronic questionnaire (eQ); and survey data exchange.



The RRM product has duly passed the security accreditation. ONS Information Assurance holds the risk management accreditation documents (RMADS) for the RRM. All future products that are built into this service will only be made 'live' when they have achieved the appropriate security accreditation.

Survey data and paradata can be extracted from the products (e.g. RRM) through reports. Reports are only available through the internal RRM site; which is only accessible by ONS staff with appropriate access.

When paradata are securely exported to Excel/SAS/r, they will be stored securely on a restricted access drive. Only the relevant project and research teams will have access. Paradata will be stored without any personal information. Two fields within the paradata set will enable researchers to link the data to their personal information: the Reporting Unit Reference number⁵ and User ID⁶. Linked personal data (from RRM) will be used for the sole purpose of re-contacting respondents for participation in the follow-up research outlined in this proposal. Paradata will only be linked to personal data if they have given prior consent to take part in future research (explained in section B6).

These linked datasets will be stored electronically in folders which have restricted access to only the ONS research team(s) working on the follow-up research.

A sample file will be created for research purposes, which contains the contact details (respondent name; business name; email address; telephone number; home/business address). This sample file will also have restricted access to only the research team(s) working on the follow-up research.

Records of incoming and outgoing calls and IDBR data are stored securely on ONS systems. Some of the records contain sensitive (i.e. identifiable) information. The research team(s) for this project will be granted access to these data at the original source and will continue to protect these data in line with the principles of the Data Protection Act (1998).

Data from follow-up research will similarly be held in a restricted-access folder (in an existing ONS drive) for the project research team(s) only.

Any research findings published will be entirely anonymous; containing no information that would identify individuals/businesses. Quotes may be used for illustrative purposes, but will never contain information which could link the data back to an individual/household/business. At present, research findings will only be disseminated throughout relevant teams within ONS.

All data will be retained and destroyed in line with the principles of the Data Protection Act (1998).

B4 Please provide details of the research protocol or methodology (e.g. data linkage, web scraping etc)
(max 500 words)

⁵ Reporting Unit Reference: The reference number which relates to the part of the business that is required to complete the MWSS.

⁶ User ID is a numeric ID that RRM assigns to the individual user / respondent. On its own it is non-identifiable.



The aim of the proposed research is to identify and follow-up on cases of interest through analysing the paradata of respondents who have completed or attempted to complete the MWSS online.

Paradata are collected through the various products that make up the service (e.g. RRM). In the future these products could change, to better meet the needs of the service. For example, an electronic questionnaire product is currently being developed, which negates the need for a hardcoded survey in RRM. New products could result in new sources of paradata.

The current research will involve analysing paradata to identify cases of interest, which suggest that a respondent has:

- a) taken a non-linear route through the survey
- b) had difficulties during completion (e.g. appearing to get stuck on certain pages; not completing/submitting the survey)

Initially, interesting cases will be identified through a process of sampling every n th case and manually reviewing the paradata associated with that case. In the future, it may be possible to develop a programme in SAS/r which would enable cases to be selected based on set criteria and associated parameters (i.e. to define what is an outlier case). This could enable the identification of a greater number of interesting cases.

When the interesting cases have been identified, their unique reference number(s), that are included in the paradata set, will be used to search for any records of any related incoming or outgoing calls on ONS databases⁷.

If all inclusion and exclusion criteria are met (e.g. the respondent has consented to being re-contacted for participation in further research *and* the business is not in enforcement for non-completion of statutory ONS surveys), the researchers will access respondents' contact details held within RRM (respondent name; business name; telephone number; e-mail address; home/business address), via their unique reference number, to invite them to take part in research to further understand their experience of completing the survey using the electronic service.

The follow-up research will be predominantly qualitative in nature and would take the format of either a semi-structured telephone or face-to-face interview. The research will examine:

- i. the reasons behind their interesting paradata
- ii. their general experience of registering for and completing the survey online
- iii. respondent needs for specific aspects of the system or survey

In addition to follow-up interviews, respondents may be invited to take part in surveys to explore (among other things) perceptions, attitudes, and experiences of using the electronic service for completing surveys. However, this method will likely be used for wider quality-assurance research with the full sample of respondents (i.e. not just those who appear to have had issues during completion).

⁷ MWSS respondents may have called ONS' Business Data Division (BDD), to ask for assistance or provide feedback while registering for the service and completing the survey. Equally, call records may relate to calls that BDD has made to MWSS respondents to chase responses and/or query data. These data provide additional context for the respondent's experience while using the RRM / eQ.



The data gathered through the follow-up research will be used to:

- i. make immediate improvements to questionnaires
- ii. make improvements to the overall service (from registration through to submission)
- iii. inform topics for further in-depth research

6.1

B5 Please outline the proposed benefits of the project (max 500 words)

Identifying issues that respondents experience while completing questionnaires means adding another data source to ONS' learning about how:

- i. questions that can be altered to capture more valid and reliable data (i.e. improve the quality of the statistics);
- ii. guidance or instructions can be altered or added to improve respondent experience and validity and reliability of data captured;
- iii. functions within an electronic questionnaire can be altered to better meet the needs of respondents;
- iv. the service (including system registration and survey completion) can be improved.

Crucially, improvements in the above areas will likely reduce the financial and time burden that is placed on respondents to complete the MWSS. This is particularly important given that for many business surveys, completion is legally enforced by ONS under the Statistics of Trade Act (1947).

In addition, carrying out this follow-up research will support ONS obligations relating to equality and diversity. Using the findings from follow-up research linked to respondent paradata, ONS can ensure that a simple, user-centric process for electronic data collection is designed to meet a broad range of accessibility needs. Users of the system and respondents to the survey will have a broad range of needs; from low levels of digital skill, to respondents with physical or mental impairments and disabilities. Ongoing research to continuously improve the way that ONS systems and processes can meet these needs is critical to maintaining a single, adaptable service that is accessible, as far as possible, for all who need to use it.

B6 Please outline the ethical issues that might arise from the proposed study and how they will be addressed (all research projects have some ethical considerations, so this section must not be left blank)

1. Issues relating to consent:

• **Issue:**

ONS will be legally compliant in regard to the collection and use of paradata. Information about paradata, and their collection and intended uses will be made "reasonably available" to respondents. The existing ONS Privacy Statement will be presented alongside the terms and conditions of the Response and Respondent Management System, which respondents are required to sign-up to prior to completing online surveys. On the advice of ONS Legal Services, an additional



statement, specific to the collection of paradata during completion of online surveys is being prepared for inclusion in the terms and conditions for the RRM. However, while ONS will meet these legal requirements, ethically, ONS has a responsibility to obtain informed consent from respondents to select them for recruitment to follow-up research, based on the analysis of their paradata.

- **Proposed solution(s):**

Section C2 (below) gives full details about how we will obtain informed consent for: (a) collecting and analysing a respondent's paradata (legal compliance); and (b) to use respondent's paradata as criteria for attempting to recruit those respondents for follow-up qualitative research.

2. Linking anonymous paradata to personal information:

- **Issue:**

Paradata sets contain no personal information, so it is not possible to identify individuals/businesses. In order to identify and select respondents for recruitment to follow-up research about their experiences, it is necessary to use the unique identifier within the paradata set, as a way of linking to the business contact details held in the RRM (respondent name; business name; email address; telephone number; address). When these data have been linked, the information is sensitive and will need to be appropriately stored and protected.

- **Proposed solution(s):**

Section B3 (above) gives full details of how we will store and protect the different datasets that form part of this project, should ethical approval be granted. All mechanisms will comply with the principles of the Data Protection Act (1998).

For overall benefits of this follow-up research, please refer to section B5 (above).

B7	How will the findings of the research be disseminated?
-----------	---

Initially, the findings from this research will be published internally to ONS, with individuals/team/departments that have an interest in or responsibility for improving the RRM system, electronic questionnaires, and the end-to-end processes of system registration and survey completion. This could be methodologists; user researchers; call centres; Survey Processing Centre; project teams involved in roll-out to electronic data collection; etc.

If there is a need, for the purposes of transparency, to publish the research findings publically, there should be no issue in doing so, as all data will be anonymised.

B8	Please outline any intended future use for products (such as linked data sets or tools) produced as a result of the research and how they will be accessed.
-----------	--



N/A - the datasets will not be converted into any products for use in other circumstances.

Section C Details of Data Subjects

6.1

C1	Data subjects to be studied
Does the Study include all subsections of the population (i.e. all ages, sex, ethnic groups etc) <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No <small>If no please detail which subsections with justification(s) below</small>	
Subsections of the population (including vulnerable groups) the project focuses on: N/A	
Justification for focusing on these subsections or groups: N/A	



C2	Please detail consent given to use data specified in section B2
----	---

Meeting legal requirements:

Respondents registering to use the Response and Respondent Management System (RRM) system to complete their electronic questionnaires are asked to sign up to the associated terms and conditions and privacy policy for the service. The current ONS Privacy Policy⁸ includes information about the capture of paradata during the use of ONS websites for the purposes of continuous improvement. On the advice of Legal Services, an additional statement will be prepared and added to the terms and conditions of the RRM, to highlight that paradata are being collected; what paradata are; and to explain what we intend to do with these data. ONS Legal Services will sign off wording and placement of this information, to ensure that it is 'reasonably available'.

Meeting ethical requirements:

In addition to meeting the legal requirements above, when respondents sign up to the terms and conditions of the system, respondents will be asked if they consent to taking part in future research to help ONS to continually improve surveys and systems for respondents. The wording of this consent is still to be determined, but it will explain that the follow-up research is to assist ONS to identify ways that the service and MWSS can be continually improved.

C3	If you are using data held by a third party please detail how you will obtain this
----	--

N/A

⁸ <http://www.ons.gov.uk/help/cookiesandprivacy>

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(16)06

The causal effect of bilingualism on academic performance

Purpose

1. The Welsh government is seeking NSDEC's advice on sharing school census data with the University of Bath. University of Bath's research will look to assess the causal link between being bilingual and academic performance.
2. The project is not sponsored by Welsh government. Welsh government's role is as a data provider and data controller.

Recommendations

3. Members of NSDEC are invited to consider the project application at **Annex A** and advise the National Statistician to:
 - i. approve the proposal and allow it to proceed;
 - ii. approve the proposal subject to minor revisions;
 - iii. recommend major revisions to the proposal and request that it be resubmitted to a future meeting once implemented; and
 - iv. reject the proposal advising that it be stopped from proceeding.

Background

4. Researchers at the University of Bath hypothesise that being a native speaker of two languages leads to higher cognitive abilities. They therefore want to analyse how bilingualism affects academic performance.
5. Previous studies have used data on immigrants to explore this relationship; however, immigrants tend to have distinct cultural and educational backgrounds which impact upon academic performance. Studying pupils in Wales may solve this identification problem.
6. Findings from the proposed research could play a role in informing language policies, such as those related to the Welsh Language Strategy. This looks to revive the Welsh language through Welsh education, which is an important cultural mission.

Adil Deedat, NSDEC Secretariat, UK Statistics Authority, 6 April 2016

List of Annexes

Annex A Application: The causal effect of bilingualism on academic performance, Aron Troth, University of Bath and Stephen Hughes, Welsh Government

Annex B Data Access Agreement template

Annex C Fair Processing Notice



National Statistician's Data Ethics Advisory Committee

Application for Ethical Review

The Application Process

This is an application form for applying for ethical review from the National Statistician's Data Ethics Advisory Committee (NSDEC). You should use the additional guidance when completing this form.

The application form should be completed in **plain English** which is understandable to lay members and all abbreviations should be explained the first time they are used. The form should contain sufficient information to ensure a thorough ethical review can take place.

Please word process the form using Arial or Times New Roman font, size 11. Where necessary expand text boxes on the form to accommodate answers, but ensure word counts are adhered to where specified.

Where sections are not relevant to your study please mark as N/A.

On completion the responsible owner should sign the application form and send to:
nsdec@statistics.gsi.gov.uk

7.1



Section A
Application Details

A1	Responsible Owner		
Full Name: [REDACTED]		Position: Lecturer	
Address: University of Bath [REDACTED]		Email: [REDACTED]	
		Telephone: [REDACTED]	
		Organisation: [REDACTED]	
<p>Declaration to be signed by the responsible owner</p> <p>I have met with and advised the applicant on the ethical aspects of this project design <i>(applicable only if the responsible owner is not the Applicant).</i></p> <p>I understand that it is a requirement for all researchers accessing the data to have undergone relevant training and to have either relevant security clearances or approved researcher status in order to access the data.</p> <p>I am satisfied that the research complies with current professional, departmental and other relevant guidelines.</p> <p>I will ensure that changes in approved research protocols are reported promptly and are not initiated without approval by the National Statistician's Data Ethics Advisory Committee.</p> <p>I will provide notification when the study is complete if it or fails to start or is abandoned.</p> <p>I will ensure that all adverse or unforeseen problems arising from the research are reported in a timely fashion to the National Statistician's Data Ethics Advisory Committee.</p> <p>I will consider all advice received from the National Statistician's Data Ethics Advisory Committee and should I be unable to implement any of the recommendations made, I will provide reasoning in writing to the Committee.</p>			
Print Name: [REDACTED] Signature: [REDACTED] Date 21 / 03 / 2016			



A2 Applicant Details (if applicant is not the responsible owner)	
Full Name	Position
Address:	Email:
	Telephone:
	Organisation:

A3 Project Information	
Project Title: The causal effect of Bilingualism on Academic Performance	
Start Date: May 2016	End Date: September 2017
Project Sponsor (select all that apply)	
<input type="checkbox"/> ONS <input type="checkbox"/> ADRN <input type="checkbox"/> GSS <input type="checkbox"/> Collaboration <input checked="" type="checkbox"/> Other University of Bath	

A4 Collaboration and Sponsors	
List of Collaborators/Sponsors	Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)
<div style="background-color: black; height: 15px; width: 100px; margin-bottom: 5px;"></div> Department of Social and Policy Sciences University of Bath	Data Access Agreement Template – Annex B

A5 Proposed Site of Research (select all that apply)

Where will the research take place?

- ☐ ONS
☐ VML
☐ HMRC Data Lab
☐ ADRC-England
☐ ADRC-Northern Ireland
☐ ADRC-Scotland
☐ ADRC-Wales
☒ Other University of Bath (data analysis), Welsh Government (data extraction)

Is this a secure site?

- ☒ Yes ☐ No

Section B

Project Details

B1 Please provide a brief high level summary of the research giving necessary background (max 500 words)

Researchers at the University of Bath (UoB) hypothesise that being a native speaker of two languages leads to higher cognitive abilities and, therefore, want to analyse how bilingualism affects academic performance. Previously, the relationship between bilingualism and academic performance has been investigated using data on immigrants. However, immigrants tend to have distinct cultural and educational backgrounds, which naturally affect academic performance. Therefore, it is hard to tell, in this context, whether the difference in immigrants' and locals' academic performance is a result of bilingualism or other factors. In other words, the effect of bilingualism couldn't be unambiguously identified in this setup.

Studying pupils in Wales could solve this identification problem. It can be argued that in



Wales's mono- and bilinguals (i.e. English and English & Welsh native speakers) share the same educational and cultural background, offering a unique opportunity to isolate the causal effect of bilingualism on academic performance.

The Knowledge and Analytical Services department at the Welsh Government (WG) hold individual data of all pupils at maintained primary and secondary schools in Wales. As all school, class, pupil, and teacher characteristics potentially affect academic performance, a comprehensive set of control variables is needed, in order to unambiguously identify the causal effect of bilingualism. Researchers at UoB wish to track the effect of bilingualism on academic performance over time, and so will examine the educational outcomes at Foundation Stage (if possible) and at Key Stages 1, 2 and 3 as well as GCSEs. All data requested are standard part of the dataset, except one that is discussed below. Names of individuals would not be required for this research (each pupil would be identified by a randomly generated number created by WG), thus minimising the risk that individual pupils could be identified.

7.1

B2	Data Use			
Type of data	Data Level			
	<i>Please specify the name of the data set</i>			
	Aggregate Data	Identifiable Data	De-identified personal data	Anonymised/ pseudo anonymised
Administrative data <i>(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)</i>	Annual school census 2005-2014: School number (anonymised) Class sizes Number of teachers Number of Welsh teachers Number of support staff Number of vacancies		Annual School Census 2005-2014: Pupil first name (coded) Pupil middle name (coded) Pupil surname (coded) Special educational needs Ethnic background National identity Free school meal entitlement Welsh language ability First	



			language Looked after indicator Month of birth Year of birth Pupil MSOA of residence Teacher assessments 2005-2014: Foundation Phase / Key Stage 1, Key Stage 2 and Key Stage 3 GCSE and A level results 2005-2014: By individual subject and grade	
Big Data <i>(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)</i>				
Survey Data <i>(please specify e.g. LFS, BRES, etc in the relevant options adjacent)</i>				
Census Data <i>(please specify year, e.g. Census 2011 in the relevant options adjacent)</i>				
Other <i>(please specify e.g. Ordnance Survey Address register in the relevant options adjacent)</i>				

B3	How will information be kept confidential and data kept secure? <i>(max 500 words)</i>
-----------	--



Any data provided by the WG will be transferred to the UoB using the WG's secure data transfer site, Data Exchange Wales Initiative (DEWi). The data will be immediately transferred to the UoB secure network, specifically to an area in which only the named contacts at UoB (and network administrators) have access. In particular, the data will be stored on a University server and in a folder to which only Áron Tóth and Matt Dickson have access. Passwords will also be applied to the information files as an additional control and also so that network administrators' access is limited to the application of necessary permissions. Further access to this data can then only be granted by provision of this password and through a request for access to the network area, which must be authorised by the responsible analyst, and separately implemented by a network administrator only on receipt of that authorisation.

Checks that the permissions applied continue to hold will be made weekly and if these checks on permissions reveal that access to additional unnamed persons has been granted, logs will be interrogated to check if there has been any unauthorised use of the information files. As soon as any such use is discovered, it will be immediately reported to the Welsh Government along with a report of how that access has been granted, whether accidentally or deliberately.

B4 Please provide details of the research protocol or methodology (e.g. data linkage, web scraping etc)
(max 500 words)

In the proposed research a statistical method, called Instrumental Variable estimation, is used to identify the causal effect of bilingualism on academic performance. This method isolates the exogenous variation in bilingualism with the aid of a so-called instrument and is thereby able to establish the causal link between bilingualism and academic performance. At the heart of this statistical method is the instrument: this is a variable that has to be uncorrelated with academic performance, while strongly correlated with bilingualism. In practice, it is usually very difficult to find a suitable instrument, i.e. a variable which meets both of these criteria. However, in the current context it appears that a variable that indicates the 'Welshness' of the pupil's name is an ideal instrument: pupils with Welsh names are naturally expected to speak Welsh with a higher probability, while the name in and of itself clearly does not affect academic performance. Indeed, an initial analysis by WG using data from the January 2015 school census revealed that 35% of pupils who have Welsh first names speak Welsh fluently, while only 10% of pupils who don't have a Welsh name speak Welsh fluently. That is, the probability of a student speaking Welsh fluently is more than three times higher if the student has a Welsh first name compared to when s/he doesn't. Thus an indicator variable for how Welsh a name is, Name Codes, appears indeed to be an ideal instrument.

Names will be coded as follows:

- i. "Welsh names" (coded 1): names that are contained in either (or both) of the following two existing books on Welsh personal names: 'Welsh First Names' by Heini Gruffudd and 'Welsh Names for your children, the complete guide' by Meic Stephens.
- ii. "Other Welsh names" (coded 2): Professor Richard Webber compiled a list of Welsh given names based on an elaborate algorithm (see e.g. [Webber: The Welsh diaspora: analysis of the geography of Welsh names](#), WG); "Other Welsh names" are names that are in the Webber list, but not in the list of "Welsh names".
- iii. "Other names" (coded 3): these are given names that are not "Welsh names" or "Other Welsh names", but they are in the list of all given names that the ONS has



published for England and Wales in 1992-2010.

- iv. “Unidentified names” (coded 0): these will be given names that cannot be matched with any names in the list provided by UoB.

Unfortunately, there is no other variable that meets both of these criteria of a valid instrument and in the current context, alternative statistical methods are unsuitable for this kind of causal analysis. Once armed with the instrument and a suitable statistical software package, it is a straightforward exercise to estimate the causal effect of bilingualism on academic performance.

Note that the research does not require the actual names of the pupils, it only requires that name codes are provided. The coding of names will be done at the WG, so researchers at UoB will receive de-identified data only. In particular, each pupil would be identified by a randomly generated number, which is uniquely created by WG for this research project.

B5 Please outline the proposed benefits of the project (max 500 words)

The UoB believe this will be the first study, which unambiguously identifies and quantifies the causal effect of bilingualism on academic performance. Identifying and measuring the effect of bilingualism on cognitive ability is important not only from a scientific perspective, but it also has important practical implications. First, many families with mixed ethnic background would naturally benefit from the results of the study. In particular, some parents often fear that raising a child bilingual would result in the child having a disadvantage in the language of education and would hinder the child’s academic progress (e.g. Hoff et al (2012): Dual language exposure and early bilingual development, Journal of Child Language, Vol 39, Issue 1). The proposed study would be in a position to unambiguously confirm or reject this concern. Second, our study would be an important contribution to the research on minority languages and language policies (e.g. Welsh Language Strategy 2012-2017, WG). For instance, the revival of the Welsh language through Welsh medium education is an important cultural mission. However, the latest poor PISA results (PISA: Programme for International Student Assessment, see <http://www.oecd.org/pisa/> for general details or the 2012 National PISA report for Wales at http://www.nfer.ac.uk/publications/PQUK02/PQUK02_home.cfm) led to some arguing that Welsh medium education may have a negative impact on students’ academic performance (e.g. Daily Mail 16/05/2015: ‘Pupils in Wales being held back...because they’re taught in Welsh’; e-Petition 09/09/2015: ‘Welsh-Medium Education: Garland or Albatross?’; Wales Online 13/02/2015: ‘Pupils taught in English are getting better grades than their Welsh-medium counterparts’). The proposed research would directly inform this policy discussion in particular, but also would speak to the wider debate on the role and importance of the revival of minority languages in general. Third, our approach will allow the distinction between learning two languages from birth and learning a second language in school, which will inform education and cultural policies. There are clear implications for education policy, for example, if learning a second language from school entry is found to have a similar impact on outcomes as learning from birth.

B6 Please outline the ethical issues that might arise from the proposed study and how they

will be addressed (all research projects have some ethical considerations, so this section must not be left blank)

The main ethical issue that may arise is the confidentiality risk related to coding of each forename, middle name and surname according to the “Welshness” of the name. The list of names and their codes have been determined following extensive consultation with academics at the Universities of Bangor, Swansea, Aberystwyth and Cardiff.

To understand how this list preserves confidentiality, please consider the following statistics: in the UoB list there are 2361 names with code 1; 2600 with code 2; and 28,842 with code 3 (naturally, UoB researchers don't know the number of unidentified names, that would be coded 0). An initial analysis by the WG revealed that in the January 2015 school census, there were 87,374 pupils with first name code 1; 9,562 with code 2; 340,399 with code 3; and 20,028 with code 0.

Please note that the researchers at UoB stand ready to consider the revision of the name categories if the WG has serious concerns that confidentiality may be at risk.

In relation to NSDEC's ethical principles:

- i. The use of data has clear benefits for users and serves the public good.
Please, see section B5.
- ii. The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately.

The name of each pupil will be coded. Each pupil will be given a random unique number that is specific to this research project. The identity of the school will also be anonymised so that the geographical location of each pupil is only identified by the MSOA of their usual residence.
- iii. The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.

There is no new technology employed in this research project. The research is conducted solely under human oversight.
- iv. Data used and methods employed are consistent with legal requirements such as the Data Protection Act, the Human Rights Act, the Statistics and Registration Service Act and the common law duty of confidence.

The sharing of individual pupil level data for research into educational achievement is lawful under [The Education \(Information About Individual Pupils\) \(Wales\) Regulations 2007](#)
- v. The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public.

All parents are provided with a Fair Processing Notice at **Annex C**

B7 How will the findings of the research be disseminated?

Results will be published in academic journals and also presented at high profile conferences, such as the Annual Conference of the Royal Economic Society (RES). Given the close media attention that these conferences receive it is expected that mainstream media would cover the results of study. (For instance, Dr Dickson has had his research covered in Daily Telegraph, Daily Mail, BBC Radio 4, Guardian, Independent, Times, New Statesman, etc). Furthermore, the researchers at UoB are expecting some invitations from policy institutions (e.g. WG) to present their results.

B8 Please outline any intended future use for products (such as linked data sets or tools) produced as a result of the research and how they will be accessed.

We do not produce products, which could be useful for further research, other than our results.

Section C

Details of Data Subjects

C1 Data subjects to be studied

Does the Study include all subsections of the population (i.e. all ages, sex, ethnic groups etc

☐ Yes ☒ No

If no please detail which subsections with justification(s) below

Subsections of the population (including vulnerable groups) the project focuses on:

All pupils in maintained schools aged 5 to 15 will be included in the study.

Justification for focusing on these subsections or groups:



C2	Please detail consent given to use data specified in section B2
-----------	--

All parents are provided with a Fair Processing Notice which follows the template below:
<http://gov.wales/docs/dcells/publications/100604s5akeepinginformeden.pdf>

The sharing of individual pupil level data for research into educational achievement is lawful under [The Education \(Information About Individual Pupils\) \(Wales\) Regulations 2007](#)

7.1

C3	If you are using data held by a third party please detail how you will obtain this
-----------	---

N/A

DAA database reference number:
Version 3.0 (January 2015)

Knowledge and Analytical Services



Llywodraeth Cymru
Welsh Government

AGREEMENT IN RESPECT OF INFORMATION PROVIDED BY THE WELSH GOVERNMENT

1. The Welsh Government has agreed to supply «OrganisationName» ("us", "we") with the information («DatasetName») described in clause (ii) of the schedule to this agreement ("the Schedule") on or by «Effective_Date» for the duration of the period set out under clause (vi) of the Schedule, subject to the terms of this agreement, and subject to the signature by us of this agreement as duly authorised signatory for and on behalf of «OrganisationName» as described in clause (i).
2. We acknowledge that the information may include personal data within the meaning of the Data Protection Act 1998 ("personal data") to which the provisions of that Act ("the DPA 1998") apply. This is set out in clause (v) of the Schedule along with any other legal considerations.
3. After receiving the information we may use it for the purpose(s) specified in clause (iii) of the Schedule, but we will not use it for any other purpose unless the Welsh Government gives us express written permission to do so. We confirm that this purpose is consistent with the aims of National Statistics.
4. We will only use the information transfer methods as set out in clause (vii) of the Schedule and will only store the information on a secure area of our network, where:
 - a. the ICT equipment used to access the network is owned and fully managed by us;
 - b. the servers or other ICT equipment used to host the network are protected by up-to-date virus-checking software; and a patching regime; with connections to other networks controlled by firewalls;
 - c. the servers (if they exist) reside in a pre-defined location within the EEA (or country included within Safe Harbour agreements) with physical and electronic access to those servers controlled and limited to known, authorised individuals only; and
 - d. the network is configured so that access to the information is restricted to the analytical contact and named individuals listed in clause (i) of the Schedule via a password controlled account.

Under no circumstances will the information be stored on a standalone device without such controls, nor will it be stored on removable media without each of hardware encryption; password protection; and separate specific agreement of the Welsh Government to this.

7.2

DAA database reference number:
Version 3.0 (January 2015)

5. We will ensure that all copies (electronic or hard) and backups of the information are managed securely and subject to the same standards as systems holding live information, again with access restricted to the analytical contact and named individuals listed in clause (i) of the Schedule.
6. We acknowledge that we have read the Code of Practice for Official Statistics: Principles 5 - Confidentiality and 8 - Frankness and accessibility¹, which describes the procedures adopted by the Welsh Government Knowledge and Analytical Services to protect the confidentiality of personal data that it holds and to comply with the provisions of the DPA 1998. We confirm that our use of the information under the terms of this agreement will be in accordance with these procedures.
7. Subject to paragraph 8, we will not allow any other person or organisation access to the information without obtaining the prior written permission of the Welsh Government and where such permission is given we will ensure that the conditions attached to such permission are met and that the permitted recipient of the information signs an agreement in respect of the information in a form approved by the Welsh Government. The conditions attached to such permission will include details of how we and the third party organisation will ensure that our ethical responsibilities and legal obligations are met during the transmission, storage, analysis, reporting on and (in due course) destruction of the information.
8. The restrictions and obligations placed on us by paragraph 7 do not apply in a situation in which we are legally obliged to disclose the information by or under legislation (for example, the Freedom of Information Act 2000), by a rule of law or by an order of a court or tribunal. Once we have a reasonable expectation that such a situation may arise we will as soon as is reasonably practicable notify the Welsh Government of that and provide the Welsh Government with such information as the Welsh Government may reasonably require in order to enable it to make representations to any person about the disclosure of the Information.
9. We will not publish any of the information or results based on analysis of the information without the prior written approval of the Welsh Government.
10. We will comply at all times with the provisions of the DPA 1998 in respect of that part of the information that is personal data (as set out in in clause (v) of the Schedule), and will not take any steps that could put at risk the confidentiality or security of the information.

¹ <http://statisticsauthority.gov.uk/assessment/code-of-practice/index.html>

DAA database reference number:
Version 3.0 (January 2015)

11. We will comply with all relevant legislation, protocols, codes of practice and ethical guidelines in respect of our use of the information, as set out in clause (iii) of the Schedule. In particular, where the information includes aggregate or anonymised data, we will not attempt to establish the identity of any individual to which the information relates.
12. If we become aware that any term of this agreement, or of any agreement entered into under paragraph 7, may have been breached, or we become aware that there may have been a breach of the DPA 1998 by any person in relation to the information, we will notify the Welsh Government as soon as is reasonably practicable as set out in paragraph 20. The means by which we will monitor and control for any potential incident is set out in clause (iv) of the Schedule.
13. We agree that the Welsh Government may terminate immediately our right to use the information under this agreement, without giving us notice, if it has reasonable grounds to believe that there may have been a breach of any term of this agreement, or of any agreement entered into under paragraph 6, or of the DPA 1998.
14. If we have reasonable grounds to believe that there may have been a breach of any agreement entered into under paragraph 7 for us to share the data with a third party, or of the DPA 1998, by any person in connection with such agreement, we will terminate immediately the right of the other party to that agreement to use the information under it.
15. If we no longer wish to use the information we may give notice to the Welsh Government advising it of that.
16. When the permitted period for our use of the information expires on «Expiry_Date» (as further detailed in clause (vi) of the Schedule), or our right to use the information is terminated by the Welsh Government, or where we have notified the Welsh Government that we no longer wish to use the information, we will at our own cost and at the Welsh Government's discretion either:
 - a. destroy or procure the destruction of all of the information in our possession or control, in a way that makes the information unrecoverable, and furnish to the Welsh Government a certificate evidencing destruction in a form acceptable to the Welsh Government; or
 - b. promptly deliver or procure the delivery of all such information to the Welsh Government in accordance with the Welsh Government's reasonable instructions.
17. We acknowledge that, in providing us with the information, the Welsh Government makes no representations and offers no guarantees as to its completeness, quality or accuracy. We also acknowledge that in no event will the Welsh Government be liable for any loss or damage including, without limitation, indirect or consequential loss or damage, arising from use or loss of use of the information.

DAA database reference number:
Version 3.0 (January 2015)

18. We acknowledge that the information is Crown Copyright, and that any reproduction, copying, broadcasting, adapting or onward supply of Crown Copyright material beyond the terms of this agreement may be a copyright infringement and will be a breach of the terms of this agreement.
19. If we are required by this agreement to give any notification to the Welsh Government, we will send that notification in writing by first class post and e-mail to the main analytical contact at the Welsh Government named in clause (i) of the Schedule.
20. We agree to fully participate in any information assurance audit or security assessment implemented by or on behalf of the Welsh Government.

DAA database reference number:
Version 3.0 (January 2015)

SCHEDULE

i. Approval

The details of the organisation receiving the information are as follows.

Organisation name	
Address	
Postcode	

As the responsible analyst from the above organisation, I approve the terms of this agreement and I agree that my organisation will abide by all the requirements specified (subject to note 1).

Name	
Position held	
Signature	
Date	
Phone	
Email	

Access to the information will be limited to a main analytical contact and the explicitly named individuals as set out below. As the main analytical contact, I agree on behalf of both myself and these named individuals that the information will neither be shared outside my organisation, nor will it be shared beyond these named individuals.

Name	
Position held	
Signature	
Date	
Phone	
Email	

Explicitly named individuals with access to the information:

Name 1	
Position held 1	
Name 2	
Position held 2	
Name 3	«OrganisationNI3Name»
Position held 3	«OrganisationNI3Pos»

DAA database reference number:
Version 3.0 (January 2015)

As the responsible analyst for the Welsh Government and Information Asset Owner in respect of this information, I authorise the provision of access to the information to the receiving organisation under the terms specified in this agreement. I delegate the responsibility to the main analytical contact as set out below for managing the provision of access to the information and to check that the terms specified in this agreement are met (subject to note 1).

Name	
Position held	
Signature	
Date	
Phone	
Email	
Organisation name	
Address	
Postcode	

As the main analytical contact for the Welsh Government in respect of this information, I accept the responsibility for managing the provision of access to the information to the receiving organisation and to check that the terms specified in this agreement are met.

Name	
Position held	
Signature	
Date	
Phone	
Email	

Note 1

The signatories believe this agreement is compliant with the statements of principle in the Code of Practice for Official Statistics ("the Code") and the specific requirements of the Principles on Confidentiality and Frankness and accessibility (Principles 5 and 8). Where this agreement may appear to contradict the statements of principle in the Code or the specific requirements of the Principles 5 and 8, the Code and the Principles 5 and 8 take precedence, unless explicitly stated.

DAA database reference number:
Version 3.0 (January 2015)

ii. Description of the information requested

iii. Purpose for which the information is requested

iv. Ensuring access to the information is controlled and limited to named contacts

v. Data Protection Act and other legal considerations

vi. Timescale

vii. Information transfer method

viii. Welsh Government contract reference (where relevant)

ix. Variable list

7.2

.

Fair Processing Notice: What the School, Local Education Authority and Government does with Information it holds on Pupils

Subject

To meet the requirements of the Data Protection Act 1998, schools are required to issue a Fair Processing Notice to pupils and/or parents summarising the information held on record about pupils, why it is held, and the third parties to whom it may be passed on.

This Fair Processing Notice provides information about the collection and processing of pupils' personal and performance information by the Welsh Assembly Government, **XXXXX** Local Education Authority (LEA) and **XXXXX** School.

The collection of personal information

The school collects information about pupils and their parents or legal guardians when they enrol at the school. The schools also collect information at other key times during the school year. Information is also received from other schools when pupils transfer.

The **School** processes the information it collects to administer the education it provides to pupils. For example:

- the provision of educational services to individuals;
- monitoring and reporting on pupils' educational progress;
- the provision of welfare, pastoral care and health services;
- the giving of support and guidance to pupils, their parents and legal guardians;
- the organisation of educational events and trips;
- planning and management of the school.

Welsh Assembly Government (WAG) & Local Education Authority (LEA)

The Welsh Assembly Government receives information on pupils normally as part of what is called the Pupil Level Annual Schools Census (PLASC). The Welsh Assembly Government uses this personal information for research (carried out in such a way that ensures individual pupils cannot be identified) and for statistical purposes, to inform, influence and improve education policy and to monitor the performance of the education service as a whole. Examples of the sort of statistics produced can be viewed at www.wales.gov.uk/statistics and www.npd-wales.gov.uk

The LEA also uses the personal information collected via PLASC to do research. It uses the results of this research to make decisions on policy and the funding of schools, to calculate the performance of schools and help them to set targets. The research is carried out in such a way that ensures individual pupils cannot be identified.

In addition WAG and LEAs receive information regarding National Curriculum assessment and Public Examination results and attendance data at pupil level..

Personal information held

The sort of personal information that will be held includes;

- personal details such as name, address, date of birth, pupil identifiers and contact details for parents and guardians;
- information on performance in internal and national assessments and examinations;
- information on the ethnic origin and national identity of pupils (this is used only to prepare summary statistical analyses);
- details about pupils' immigration status (this is used only to prepare summary statistical analyses);
- medical information needed to keep pupils safe while in the care of the school;
- information on attendance and any disciplinary action taken;
- information about the involvement of social services with individual pupils where this is needed for the care of the pupil.

Organisations who may share personal information

Information held by the School, LEA and the Welsh Assembly Government on pupils, their parents or legal guardians may also be shared with other organisations when the law allows, for example with;

- other education and training bodies, including schools, when pupils are applying for courses, training, school transfer or seeking guidance on opportunities;
- bodies doing research for the Welsh Assembly Government, LEA and schools, so long as steps are taken to keep the information secure;
- central and local government for the planning and provision of educational services;
- social services and other health and welfare organisations where there is a need to share information to protect and support individual pupils;
- various regulatory bodies, such as ombudsmen and inspection authorities, where the law requires that information be passed on so that they can do their work.

Pupils have certain rights under the Data Protection Act, including a general right to be given access to personal data held about them by any "data controller". The presumption is that by the age of 12 a child has sufficient maturity to understand their rights and to make an access request themselves if they wish. A parent would normally be expected to make a request of child's behalf if the child is younger.

If you wish to access your personal data, or that of your child, then please contact the relevant organisation in writing. Details of these organisations can be found on the following website (**School or LEA website**) or for those pupils/parents where this is not practical, a hard copy can be obtained from the school (**add details of who should be contacted**)

Other information

The Welsh Assembly Government, LEA and school place a high value on the importance of information security and have a number of procedures in place to minimise the possibility of a compromise in data security.

The Welsh Assembly Government, LEA and School will endeavour to ensure that information is kept accurate at all times. Personal information will not be sent outside the United Kingdom.

Your rights under the Data Protection Act 1998

The Data Protection Act 1998 gives individuals certain rights in respect of personal information held on them by any organisation. These rights include;

- the right to ask for and receive copies of the personal information held on **you**, although some information can sometimes be legitimately withheld;
- the right, in some circumstances, to prevent the processing of personal information if doing so will cause damage or distress;
- the right to ask for wrong information to be put right;
- the right to seek compensation if an organisation does not comply with the Data Protection Act 1998 and you person suffer damage;
- in some circumstances a pupil's parent or legal guardian *may* have a right to receive a copy of personal data held about a pupil in their legal care. Such cases will be considered on an individual basis where the individual is deemed to have insufficient understanding of their rights under the Act.

You also have the right to ask the Information Commissioner, who enforces and oversees the Data Protection Act 1998, to assess whether or not the processing of personal information is likely to comply with the provisions of the Act.

Seeking further information

For further information about the personal information collected and its' use, if you have concerns about the accuracy of personal information, or wish to exercise your rights under the Data Protection Act 1998, you should contact;

- your child's school on **XXXX XXXXXX**;
- your LEA on **XXXX XXXXXX**;
- the Welsh Assembly Government's data protection officer at, The Welsh Assembly Government, Cathays Park, Cardiff, CF10 3NQ;
- the Information Commissioner's office help line can be contacted on 01625 545 745;
- information is also available from www.ico.gov.uk

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(16)07

Education identifiers - read through

Purpose

1. This Scottish government seeks to develop an approach to data linkage projects, which further reduces privacy risks and improves efficiency. The proposal presented at **Annex A** looks to implement this approach for education data and aims to do this in a way that ensures that governance controls remain robust and, in particular, does not undermine the oversight data controllers have over use of data.

Recommendations

2. Members of NSDEC are invited to advise the National Statistician to:
 - i. approve the proposal and allow it to proceed;
 - ii. approve the proposal subject to minor revisions;
 - iii. recommend major revisions to the proposal and request that it be resubmitted to a future meeting once implemented; or
 - iv. reject the proposal advising that it should be stopped from proceeding.

Background

3. The Education Analytical Services (EAS) within Scottish government provides a range of analytical and intelligence services covering three directorates in DG Learning and Justice – Children and Families, Learning and Employability, Skills and Lifelong Learning.
4. The EAS takes a leading role in data linkage supporting the efficient use of education data for linkage purposes within controlled and secured environments.
5. Researchers can apply to EAS for access to education data for linkage based projects. The effective and appropriate use of data is considered by EAS Division to be an important aspect of the work of the Division, and as such EAS Division looks to facilitate the sharing of data with the education and wider communities.
6. Decisions on whether to enable access to data held by EAS Division are ultimately the responsibility of the Head of the Division, but advice is provided by the Divisional Data Access Officer and the EAS Data Access Panel (DAP).
7. ScotXed is a unit within EAS. ScotXed supports and develops a significant number of secure, efficient and effective electronic data exchanges between partners in the Scottish government and wider service communities.
8. ScotXed provide data to our analytical colleagues within the Scottish government for national and international statistical publications. Analytical products contribute to the evidence base for policy development and making. We adhere to the principles of the Code of Practice for Official Statistics and the Data Protection Act (1998) when collecting data for research and statistical purposes.
9. ScotXed provides the secretariat for the DAP. All applications are reviewed in relation to the privacy risks and public benefits around: the risks associated with the data requested; the accreditation of researchers; security; likelihood and impact of a breach; public and policy interest.

10. National Records of Scotland (NRS) is a non-ministerial department of the Scottish government with the purpose to collect, preserve and produce information about Scotland's people and history and make it available to inform current and future generations.
11. The NRS Indexing Team is an integral part of the new Scottish Informatics and Linkage Collaboration (SILC). SILC is a Scotland-wide collaboration between NHS National Services Scotland (NSS), academic partners in the Farr Institute, the Administrative Data Research Centre and NRS, which was developed following consultation in 2012 on the aims, benefits and challenges to data linking.
12. Currently the linkage of data is repeated for each project on a project by project basis. This means that personal identifiers are repeatedly shared and linked by indexing teams. **Annex A** sets out a proposal, which would allow for the development of an artificially created reference number, concealing the identity of individuals, and which can be shared with the indexing team when they are required to facilitate an education data linkage project.

Albert King, Head of ScotXed, Scottish Government, 8 April 2016

List of Annexes

Annex A Application: Education Identifiers - read through, Mr Albert King

Annex B Memorandum of understanding

Annex C Education read through diagram



National Statistician's Data Ethics Advisory Committee

Application for Ethical Review

The Application Process

This is an application form for applying for ethical review from the National Statistician's Data Ethics Advisory Committee (NSDEC). You should use the additional guidance when completing this form.

The application form should be completed in **plain English** which is understandable to lay members and all abbreviations should be explained the first time they are used. The form should contain sufficient information to ensure a thorough ethical review can take place.

Please word process the form using Arial or Times New Roman font, size 11. Where necessary expand text boxes on the form to accommodate answers, but ensure word counts are adhered to where specified.

Where sections are not relevant to your study please mark as N/A.

On completion the responsible owner should sign the application form and send to:
nsdec@statistics.gsi.gov.uk

8.1



Section A
Application Details

A1	Responsible Owner		
Full Name: [REDACTED]		Position: [REDACTED]	
Address: ScotXed, [REDACTED] [REDACTED]		Email: [REDACTED]	
		Telephone: [REDACTED]	
		Organisation: Scottish Government	
<p>Declaration to be signed by the responsible owner</p> <p>I have met with and advised the applicant on the ethical aspects of this project design <i>(applicable only if the responsible owner is not the Applicant)</i>.</p> <p>I understand that it is a requirement for all researchers accessing the data to have undergone relevant training and to have either relevant security clearances or approved researcher status in order to access the data.</p> <p>I am satisfied that the research complies with current professional, departmental and other relevant guidelines.</p> <p>I will ensure that changes in approved research protocols are reported promptly and are not initiated without approval by the National Statistician's Data Ethics Advisory Committee.</p> <p>I will provide notification when the study is complete if it or fails to start or is abandoned.</p> <p>I will ensure that all adverse or unforeseen problems arising from the research are reported in a timely fashion to the National Statistician's Data Ethics Advisory Committee.</p> <p>I will consider all advice received from the National Statistician's Data Ethics Advisory Committee and should I be unable to implement any of the recommendations made, I will provide reasoning in writing to the Committee.</p>			
Print Name: [REDACTED] Signature: [REDACTED] Date: 8 April 2016			



A2 Applicant Details (if applicant is not the responsible owner)	
Full Name	Position
Address:	Email:
	Telephone:
	Organisation:

8.1

A3 Project Information	
Project Title: Education Identifiers Read-through	
Start Date:	End Date: Continuous
Project Sponsor (select all that apply)	
<input type="checkbox"/> ONS <input type="checkbox"/> ADRN <input type="checkbox"/> GSS <input type="checkbox"/> Collaboration <input checked="" type="checkbox"/> Other (Please specify). Scottish Government Analytical Leadership Group	

A4 Collaboration and Sponsors	
List of Collaborators/Sponsors	Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)
Education Analytical Services, ScotXed Unit, Scottish Government National Records of Scotland (NRS)	Annex B Memorandum of Understanding



A5	Proposed Site of Research (select all that apply)
<p>Where will the research take place?</p> <p><input type="checkbox"/> ONS</p> <p><input type="checkbox"/> VML</p> <p><input type="checkbox"/> HMRC Data Lab</p> <p><input type="checkbox"/> ADRC-England</p> <p><input type="checkbox"/> ADRC-Northern Ireland</p> <p><input type="checkbox"/> ADRC-Scotland</p> <p><input type="checkbox"/> ADRC-Wales</p> <p><input checked="" type="checkbox"/> Other (please specify) National Records of Scotland</p>	
<p>Is this a secure site?</p> <p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p>	

Section B

Project Details

B1	Please provide a brief high level summary of the research giving necessary background <i>(max 500 words)</i>
<p>Education Analytical Services (EAS) has taken a leading role in data linkage, supporting the legal, ethical and efficient use of education data for linkage purposes within controlled and secure environments. This has led to increased interest from academia and other areas of government in access to education data for linkage and formal applications for projects. The health sector has been particularly active in bringing forward research projects which produce evidence to inform clinical practice with a better understanding of the wider outcomes of health interventions.</p> <p>EAS ensures that each applicant seeking to undertake a data linkage project ('education linkage project') works with recognised linkage services, such as the electronic Data Research and Innovation Service (eDRIS), Scottish Informatics and Linkage Collaboration</p>	

(SILC) or the Administrative Data Research Network (ADRN). However, this still requires the linkage process to be repeated for every project on a project by project basis. As a result, personal identifiers must be repeatedly shared with and linked by the indexing teams. Our project seeks to improve the efficiency and privacy protection of the process by establishing an index¹ between education identifiers and the indexing 'spine'² (referred to as 'Read Through'). This will be an artificially created reference number that conceals the identity of the person, meaning that no personal identifiers will need to be sent to the indexing team when they are required to perform an education linkage project. It should be noted that NRS are the Trusted Third Party for Scottish Government linkage and, in line with the standard separation of functions model, solely receive personal identifiers for the purpose of providing an indexing and linkage service, at no point do NRS ever receive attribute data. This separation of functions ensures that those handling the identifiable data do not see the attribute data, and vice versa.

The project seeks to develop a process to create the index key that meets both technical and governance criteria. It is a critical success criterion of the governance model we will develop for our project that data controllers (EAS in the case of education data) remain in control of their data and decisions about how it will be used.

We anticipate that the process in outline will involve:

- i. Following completion of pupil census, education identifiers would be indexed to the 'spine'. The indexing spine³ is a thin list of the whole population of Scotland including limited demographic information to allow individuals to be matched to the spine. The spine allows us to link individual sources to the spine without requiring all datasets in a project to be held at the same time and also allows us the ability to carry out 'read through'
- ii. ScotXed (part of EAS) and NRS would then hold an index key, which has no personal data attached to it. The index key would be used only for statistical and research data linkage and is not a persistent identifier used in any live or operational systems.
- iii. On a project by project basis, project specific index-ids would be generated by EAS and sent to the linkage team in NRS. NRS could then use the index keys to generate new project indexes to be used in the individual linkage projects
- iv. This will allow EAS to retain appropriate governance and controls on use of the data.
- v. It will reduce the frequency with which we need to supply and transmit personal identifiers to linkers and thus enhance privacy for individuals and reduce the burden on the NRS Indexing team.

¹ [Index](#): A dataset of individual references that can be used for cross-referencing between sources.

² [Spine or Population Spine](#): A dataset created by cross-referencing indices that allows for an accurate count of the population.

³ The NRS indexing spine comprises data from birth registrations, death registrations and GP registrations in Scotland. It is maintained on an on-going basis.

B2	Data Use																																					
	<table border="1"> <thead> <tr> <th rowspan="2">Type of data</th> <th colspan="4">Data Level <i>Please specify the name of the data set</i></th> </tr> <tr> <th>Aggregate Data</th> <th>Identifiable Data</th> <th>De-identified personal data</th> <th>Anonymised/ pseudo anonymised</th> </tr> </thead> <tbody> <tr> <td>Administrative data <i>(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)</i></td> <td></td> <td>Pupil census from 2007 and beyond</td> <td></td> <td></td> </tr> <tr> <td>Big Data <i>(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)</i></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Survey Data <i>(please specify e.g. LFS, BRES, etc in the relevant options adjacent)</i></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Census Data <i>(please specify year, e.g. Census 2011 in the relevant options adjacent)</i></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Other <i>(please specify e.g. Ordnance Survey Address register in the relevant options adjacent)</i></td> <td></td> <td>NRS Linkage Spine</td> <td></td> <td></td> </tr> </tbody> </table>				Type of data	Data Level <i>Please specify the name of the data set</i>				Aggregate Data	Identifiable Data	De-identified personal data	Anonymised/ pseudo anonymised	Administrative data <i>(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)</i>		Pupil census from 2007 and beyond			Big Data <i>(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)</i>					Survey Data <i>(please specify e.g. LFS, BRES, etc in the relevant options adjacent)</i>					Census Data <i>(please specify year, e.g. Census 2011 in the relevant options adjacent)</i>					Other <i>(please specify e.g. Ordnance Survey Address register in the relevant options adjacent)</i>		NRS Linkage Spine		
Type of data	Data Level <i>Please specify the name of the data set</i>																																					
	Aggregate Data	Identifiable Data	De-identified personal data	Anonymised/ pseudo anonymised																																		
Administrative data <i>(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)</i>		Pupil census from 2007 and beyond																																				
Big Data <i>(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)</i>																																						
Survey Data <i>(please specify e.g. LFS, BRES, etc in the relevant options adjacent)</i>																																						
Census Data <i>(please specify year, e.g. Census 2011 in the relevant options adjacent)</i>																																						
Other <i>(please specify e.g. Ordnance Survey Address register in the relevant options adjacent)</i>		NRS Linkage Spine																																				

B3	How will information be kept confidential and data kept secure? <i>(max 500 words)</i>
<p>The process to develop the index key will follow the Guiding Principles for Data Linkage [see link], these are a set of principles that have been developed by Scottish government to support the legal, ethical and efficient use of data for linkage purposes within a controlled and secure environment. A Memorandum of Understanding will be put in place between EAS and NRS to define roles and responsibilities. NRS uses a secure data transfer platform called 'Thru' to transfer data.</p> <p>The resulting index key will be held by NRS and EAS in existing secure settings. EAS will hold the linkage indexes in a restricted and secure data warehouse on the Scottish Government network separately from our main analytical datasets. Scottish Government and NRS are ISO27001 compliant organisations. ISO27001 is a specification for an information security management system that includes legal, physical and technical controls. Only those</p>	



involved in administration of IT systems and data linkage within the ScotXed unit and NRS will have access to these indexes. NRS will hold indexes within the secure environment established to support SILC. The indexing environment is a strictly controlled isolated environment within NRS.

Only Scotxed will have the ability to link the index back to actual pupil data and only NRS will have the ability to link the Index back to the spine. Therefore, the indexes and no other body will ever see these indexes so they are useless to any other organisation.

B4 Please provide details of the research protocol or methodology (e.g. data linkage, web scraping etc)
(max 500 words)

The linkage will create both a read through between personal identifiers documented processes. It will ensure that existing governance processes have appropriate control over use of the new index. These governance processes will ensure that for specific projects appropriate review is carried out in line with existing data access protocols and governance structures. It will use established SILC processes to ensure separation of personal identifiers⁴ from attribute data⁵. By establishing a read through, the initial phases of that process will be simplified.

One of the deliverables of the project will be to document in detail the process which will be designed to reduce privacy risks and simplify future linkage projects.

No attribute data will be shared in order to establish the linkage to the spine. Analytical projects will be required to make a formal application which will be reviewed by the EAS Data Access Panel. Attribute data will not be shared with the indexing team. Please refer to the attached diagram for an illustration of the anticipated process and movement of data.

We anticipate that the process in outline will involve:

- i. Following completion of pupil census, education identifiers would be indexed to the 'spine'.
- ii. ScotXed and NRS would then hold an index key.
- iii. On a project by project basis, project specific index-ids would be generated and sent to the linkage team.
- iv. This will allow EAS to retain appropriate governance and controls on use of the data.

B5 Please outline the proposed benefits of the project (max 500 words)

⁴ [Personal Identifiers / Identifiable data](#): Information about a living individual who can be identified from that information and other information which is in the data controller's possession. In our project comprises includes date of birth, gender, postcode.

⁵ Attribute Data: The data recorded about an individual which cannot be used easily to identify them. In our project this comprises education data collected through the pupil census, attainment data, school attendance data.



We anticipate that the project will deliver the following benefits:

- i. Reduced privacy risks as individual identifiers required for indexing would be shared once rather than on a project-by-project basis
- ii. Increased use of EAS data for research with benefits to public policy and academic research outputs that inform practice in education, health and other fields
- iii. Reduced burden on EAS as identifiers would need to be extracted only once
- iv. Improved efficiency of linkage as the indexing team would need to carry out indexing only once

Since EAS has developed the use of education data for data linkage studies, a number of research projects have been conducted. These include published papers on the educational outcomes of mode of birth delivery⁶, projects looking at the dental health of looked after children, month of conception and later life learning disabilities and access to education.

B6 Please outline the ethical issues that might arise from the proposed study and how they will be addressed (*all research projects have some ethical considerations, so this section must not be left blank*)

It is important to note that the scope of this data sharing for linkage will be limited to personal identifiers (postcode, gender, date of birth) required to establish the linkage to the spine. Further data sharing required for analytical projects will be subject to formal applications and review by governance groups such as the EAS Data Access Panel.

Principle i - Our project will deliver benefits to data subjects by further reducing privacy risks. Additionally, it will provide benefits to data subjects and the wider public by improving the efficiency of linkage projects. These linkage projects serve the public good by providing outputs which inform practice and policy making. The improvement in efficiency will increase capacity to support these projects and reduce the time taken for them to produce outputs.

Principle ii -Our project seeks to protect individuals identity through careful separation of roles, by ensuring that appropriate technical and physical controls are in place and by ensuring appropriate governance and oversight are established. Data subjects are kept informed about how their data is used as outlined in (C2).

Principle iii - Our project will use proven tools and technologies for the indexing process. Part of the work of the indexing team is the continuous review and enhancement of their tools to ensure that they remain fit for purpose. The team is involved in developmental work to explore new methods of indexing which may enhance privacy preserving linkage.

As much as they present opportunities, technical developments can also create risks. One

⁶ Educational outcomes following breech delivery: a record-linkage study of 456 947 children; International Journal of Epidemiology, Novel cross-sectoral linkage of routine health and education data at an all-Scotland level: a feasibility study. Lancet 2013



example is the risk of re-identification of data subjects from published outputs. Although our project will not publish any data about individuals it is important to note that the governance structures which review new linkage projects consider these risks as part of their evaluation of each project.

Principle iv - The information given in (C2) sets out the legal basis on which the data are collected and may be shared with NRS for these purposes.

Principle v - Although our project does not directly produce research outputs, the [Scottish Government consultation](#) on public acceptability on the establishment of SILC addresses issues of relevance and in particular the issue of developing read through arrangements for key data set such as education.

Principle vi - Our privacy notice the 'School Handbook Insert' is published on the Scottish Government website. Additionally, data subjects or their guardians are provided with this information through their local education authority. The notice states that data may be shared with NRS and academic institutions for research purposes.

Further considerations may include:

- The risk, or perceived risk that a single database of personal records is being created
- The risk that data controllers lose control over decisions about how data are processed in line with their obligations to do so legally and fairly

We have sought to address these issues directly through the design of the process. In particular, data controllers will retain control over how their data is used. This is ensured both through governance structures (NRS will be processors on behalf of ScotXed who will remain data controllers) and by ensuring that data required for research is not exchanged until individual research projects achieve approval under established governance regimes.

8.1

B7	How will the findings of the research be disseminated?
-----------	---

This work is not intended to lead to published papers. However, the findings of this work will be disseminated through the SILC networks and ADRN. Additionally, it is anticipated that this work will lead to further publication of research papers produced by projects enabled by the established read through.

B8	Please outline any intended future use for products (such as linked data sets or tools) produced as a result of the research and how they will be accessed.
-----------	--

The principle outputs of this work will be an established read-through between education identifiers and the SILC spine and documented processes for maintaining and governing the read through. These products will be available within the SILC, Farr and ADRN networks and accessible through the SILC linkage mechanism.



Each of these routes requires governance approvals before it may proceed. In the case of ADRN, for example, this includes review and approval through ethical panels and the [ADRN approvals panel](#). In the case of SILC projects, approval is required from the Public Benefits & Privacy Panel. Furthermore, all projects require review and approval from the EAS Data Access Panel.

Section C

Details of Data Subjects

C1	Data subjects to be studied
<p>Does the Study include all subsections of the population (i.e. all ages, sex, ethnic groups) etc</p> <p><input type="checkbox"/> Yes <input checked="" type="checkbox"/> No <i>If no please detail which subsections with justification(s) below</i></p>	
<p>Subsections of the population (including vulnerable groups) the project focuses on:</p> <p>The project applies to all subjects within the scope of the Scottish Government school pupil census.</p>	
<p>Justification for focusing on these subsections or groups:</p> <p>Proposals focusing on linkage of education data.</p>	

C2	Please detail consent given to use data specified in section B2
-----------	--

EAS ensure that data subjects are informed about how their data may be used and that appropriate legal gateways exist to enable data sharing. The sharing of this data is necessary for the administration of the functions of government. The Scottish Government is a producer of Official Statistics and bound by the Statistics and Registration Services Act 2007, which establishes the Code of Practice (CoP) for Official Statistics (section 10). This requires us to:

- Meet user needs, as defined under principle 1 of the CoP, including: dissemination of official statistics to meet the requirements of informed decision making by government, public services, business, researchers and the public and to maximise public value.
- Make statistics available to all users, as in principle 8 of the CoP, including: make statistics available in as much detail as is reliable and practicable, subject to legal and confidentiality constraints and ensure that official statistics are disseminated in forms that enable and encourage analysis and re-use.

The Education (Scotland) Act 1980 and the Education (Schools) Act 1992 provide the legislation to collect data about schools and pupils (including information about the continuing education of pupils leaving a school, or the employment or training taken up by such pupils on leaving).

The provision of data by EAS to others is communicated to subjects through the ScotXed School Handbook inserts. They inform pupils and parents about how the data the Scottish Government and its Local Government partners collect about pupils will be used, why it is needed and what we do to protect the information. This clarifies that individual data is used for statistical and research purposes only. The handbook insert notes that information (including at the individual level) may be shared with partners, including Education Scotland and National Records of Scotland.

<http://www.gov.scot/Topics/Statistics/ScotXed/SchoolEducation/SchoolPupilCensus/SchoolHandbookInsertpupils>

<http://www.legislation.gov.uk/ssi/2012/130/made>

<http://www.scotland.gov.uk/Publications/2012/09/8694>

Data will not be transferred outside the European Economic Area unless the country or territory to which the data are to be transferred provides an adequate level of protection for personal data (e.g. one of the European commission recognised non-EEA countries which are deemed to provide an adequate level of protection or a company in the United States which has signed up to the 'Safe Harbour' agreement).

Scottish Government, NRS and NHS Scotland carried out a formal public consultation on the design of the national data linkage centre (SILC) in 2013. This consultation included the proposal to develop read through arrangements for key data set such as education. There were no significant objections to the proposal in the consultation. For more information on public acceptability research, see:

<http://www.gov.scot/Topics/Statistics/datalinkageframework/PublicAcceptabilityResearch>



C3	If you are using data held by a third party please detail how you will obtain this
N/A	

**Data Controller/Data Processor Memorandum of
Understanding (MoU)
Service Requirements and Obligations
Between
The Scottish Government
and
National Records of Scotland
For the purpose of providing Indexing Services for
Education Analytical Services Statistical Data**

Contents

1. The Agreement	4
1.1. Statement of Intent.....	4
1.2. Parties to the Agreement	4
1.3. Terms of Agreement	4
1.4. Obligations	4
1.5. Document Sign-Off and Security.....	4
2. The Service	5
2.1. Services in Scope and Description.....	5
3. Service Governance & Compliance	5
3.1. Governance Activities	5
3.2. Compliance	6
3.3. Requirements.....	6
3.4. Data	7
3.5. Privacy and Confidentiality	7
3.6. Personal Data	7
3.7. Loss of Data	8
3.8. Ownership of Data	9
3.9. Intellectual Property Rights	9
3.10. Records Retention	9
3.12. Audit Rights	10
3.13. Quality Assurance.....	10
3.14. Constraints.....	10
4. Termination and Exit.....	10

4.1. Termination Rights	10
4.2. Termination Process and Responsibilities	11
5. Charging and Payments	11
6. Signatories.....	11

1. The Agreement

1.1. ***Statement of Intent***

- 1.1.1. This document (“the Agreement”) is a memorandum of understanding between the National Records of Scotland (“the Provider”) and the Scottish Government (“the Customer”) in relation to the provision of data processing (including indexing) services by the Provider to the Customer and the legal requirements that the Provider must meet.

1.2. ***Parties to the Agreement***

- 1.2.1. **The Registrar General of Births, Deaths and Marriages for Scotland** (the Provider) having its headquarters at HM GENERAL, Register House, 2 Princes Street, Edinburgh EH1 3YY, and
- 1.2.2. **The Scottish Government** (the Customer) having its headquarters at St Andrew’s House, Regent Road, Edinburgh. EH1 3DG

1.3. ***Terms of Agreement***

- 1.3.1. The term of the Agreement, which covers the services specified in Section 2.1, will commence on 01/03/16. The Agreement will be reviewed annually by the parties.

1.4. ***Obligations***

- 1.4.1. The Provider will deliver the agreed services within the agreed timeframes, standards and costs, and in accordance with legal obligations which the Customer requires the Provider to meet.
- 1.4.2. Both parties agree to act in good faith and in a reasonable and timely manner with regard to the operation of the Agreement.
- 1.4.3. The parties agree to meet the costs, where material, of rework required as a result of their respective actions/errors following senior level consultation where appropriate.

1.5. ***Document Sign-Off and Security***

- 1.5.1. The Agreement is to be signed by the Information Asset Owner (IAO) in duplicate and retained by both parties in a secure format. Electronic versions of the Agreement held by either party must also be kept in a secure format.

2. The Service

2.1. *Services in Scope and Description*

2.1.1. The Provider will provide the following services

- a) **Advisory Service** relative to linkage, indexing and data in general, including information on the Provider's data sets and associated variables where appropriate; the indexed file (data controllers ID and IndexID) and necessary information on the master indexing file (UID and IndexIDs); the overall quality of the data provided by the Customer to allow for provision of the Services by the Provider; and co-producing a set of access rules and protocols for the handling of Personal Data as part of the Service with the Customer and those across the broader Scottish Informatics and Linkage Collaboration (SILC). These rules and protocols will cover access to information and systems for the duration of this MoU, and may include procedures for the withdrawal of access and the return or deletion (at the Customer's discretion) of all Personal Data on completion of the relevant Services and/or termination of this Agreement.
- b) **Indexing Service** as part of the 'separation of functions' approach being used during the linkage process, in-line with the Data Linkage Framework and Scottish Informatics and Linkage Collaboration (SILC). This will include assigning a non-personal unique identifier (UID) to the Personal Data or local identifiers provided by the Customer via secure file transfer, together with such other information as the Provider may reasonably require in order to provide the Services; maintaining 'look-up tables' and 'read-through' which will allow for the development of more effective and efficient Service Provision; standard formatting of data before matching to ensure high quality accurate outputs as a result of the linkage being undertaken.
- c) **Secure Data Transfer Service** as agreed with the Customer to allow for the secure transfer of Personal Data and attached non-personal UID's between the Provider and the Customer; to securely transfer Index Keys to the Linkage Provider (NHS National Services Scotland or others as advised by the Customer).

3. Service Governance & Compliance

3.1. *Governance Activities*

- 3.1.1. The Provider and the Customer will operate under their established governance structures relative to their respective organisations.

Additionally, both the Provider and the Customer will be governed in relation to the processing of data pursuant to this agreement by the Scottish Informatics and Linkage Centre – Strategic Management Board (SILC-SMB).

3.2. Compliance

3.2.1. As a minimum, the Provider shall comply in all respects and shall at all times act in such a manner to assist the Customer to comply, with the following:

- a) the Data Protection Act 1998 (DPA), and all codes and guidance issued pursuant thereto;
- b) the Human Rights Act 1998;
- c) The Freedom of Information (Scotland) Act 2002 and Environmental Information (Scotland) Regulations 2004
- d) the common law duty of confidentiality;
- e) the Scottish Government Identity Management and Privacy Principles;
- f) the Scottish Government's Joined Up Data for Better Decisions Strategy;
- g) the Scottish Government's Guiding Principles for Data Linkage; and
- h) The Scottish Government's Security Classifications Policy and Handling Guidance in relation to:
 - a. Transferring Personal Data securely
 - b. Communication of personal data by: post, fax, phone, email and other electronic transfers of data;
 - c. Storage of personal data on mobile devices including laptops, USB memory sticks, PDAs, Blackberries or any other mobile device or media such as smart phones, CD or DVD; and
 - d. Anonymisation of data whenever feasible or appropriate;
- i) The Provider shall notify the Customer immediately if it becomes aware of any unauthorised or unlawful Processing, damage to or destruction of the Data, or if such Data becomes damaged, corrupted or unusable. The Provider shall follow the Procedure as detailed in Section 3.7.

3.3. Requirements

3.3.1. The Provider undertakes that it shall process the Personal Data strictly in accordance with the Customer's instructions for the processing of that Personal Data.

- 3.3.2. The Provider will process the Personal Data for the purpose of providing the Services (Section 2.1).
- 3.3.3. The Customer will transfer the Personal Data in-line with the Government Security Classification Policy (Section 3.2).
- 3.3.4. The Provider agrees to inform the Customer as soon as possible (ideally within 3 working days) of all subject access requests which may be received from the Data Subjects. Where necessary the Provider will assist the Customer in processing the requests in line with the requirements of the DPA.

3.4. Data

- 3.4.1. The Customer shall provide the relevant Personal Data (see Section 3.6) to the Provider together with such other information as the Provider may reasonably require in order to provide the Indexing Service. All relevant Personal Data (see Section 3.6) will be transferred to the Provider in-line with the Customer's security requirements.
- 3.4.2. The Customer shall provide only the minimum data necessary for the provision of the Services.
- 3.4.3. The instructions given by the Customer to the Provider in respect of the Personal Data shall at all times be in accordance with the laws of the United Kingdom.

3.5. Privacy and Confidentiality

- 3.5.1. The Provider will treat the Personal Data, and any other information provided by the Customer as confidential, and will ensure that access to the Contract Personal Data is limited to only those employees who require access to it for the purpose of the Provider carrying out the Services and complying with this Agreement and the Provider will ensure that all such employees have undergone training in the law of data protection, their duty of confidentiality under contract and in the care and handling of Personal Data.
- 3.5.2. The Provider will not disclose the Contract Personal Data to a third party under any circumstances other than at the specific written request of the customer, unless the disclosure is required by law.

3.6. Personal Data

- 3.6.1. The scope and type of Personal Data that may be provided to or accessed by the Provider as part of the Services consists of:-
 - Local unique identifiers;
 - Forename, surname and any other name fields
 - Gender
 - Date of Birth

- address;
- postcode;
- Unique Property Reference Number (UPRN);
- Unique Delivery Point Reference Number (UDPRN).

- 3.6.2. The Provider undertakes (on its own behalf and on behalf of the Customer) to treat all Personal Data in accordance with the provisions and principles of the DPA and to ensure only those of its staff who require to access Personal Data in the performance of their duties under this Agreement are able to do so, and that such staff are appropriately trained as a safe researcher and vetted to ensure their reliability.
- 3.6.3. The Provider shall only access Personal Data which is directly relevant to the effective execution by it of the terms of the Services.
- 3.6.4. The Provider will employ appropriate operational and technological processes and procedures to keep the Personal Data safe from unauthorised use or access, alteration, transmission, publication, loss, destruction, theft or disclosure. Such technical and organisational measures will follow the Customer's Security Classification Policy and Handling Guidance (Section 3.2).
- 3.6.5. The Provider's organisational, operational and technological processes and procedures adopted are required to comply with the requirements of the UK HMG Security Policy Framework and UK HMG Information Security policies, guidelines and standards, including those produced by the CESG
- 3.6.6. The Provider will not keep the Personal Data on any laptop or other removable drive or device unless that device is protected by being fully encrypted to the standards required under the Customer's Mobile Data Protection Standard available within the Scottish Government's Security Classifications Policy and Handling Guidance (Section 3.2) and the use of the device or laptop is necessary for the provision of the Services. Where this is necessary, the Provider will keep an audit trail of all such laptops, drives, and devices that Contract Personal Data is held on.

3.7. *Loss of Data*

- 3.7.1. The Provider shall notify the Customer immediately if it becomes aware of any loss of Personal Data. Following this, the Customer, as the Data Controller, will agree with the Provider the decisions to be taken and management of the incident. The Provider shall follow the Customer's Security Incident Reporting Policy by reporting, documenting and following up on all incidents, and reporting to senior management in both parties immediately. The outcome of any such investigation under the Security Incident Reporting Policy shall be notified to the Customer. The Customer will support the Provider to meet these obligations as required.

3.8. Ownership of Data

- 3.8.1. All data remains the ownership of the Customer and other data controllers involved (either alone or jointly or in common with other persons). At no point will the Provider own the data.
- 3.8.2. The Provider will not sub-contract any of the Processing Activity or Data without the explicit written consent of the Customer. Where such consent is provided the Provider will ensure that any sub-Data Processor it uses to Process the Contract Personal Data enters into a written agreement which imposes the same obligations as are imposed on the Processor under this Agreement.
- 3.8.3. Where the sub-Data Processor fails to fulfil its obligations under any sub-Processing agreement, the Processor shall remain fully liable to the Customer for the fulfilment of its obligations under this Agreement.

3.9. Intellectual Property Rights

- 3.9.1. The Provider has no intellectual property rights pertaining to the processed data. All Intellectual Property Rights remain with the Customer and other data controller(s) involved (either alone or jointly or in common with other persons) unless stipulated elsewhere (for example in the project proposal).
- 3.9.2. The Provider shall on reasonable and justifiable request share with the Customer such Intellectual Property as is created or employed by the Provider in order to process the data for the purpose of providing the Services including business processes, computer code and algorithms.

3.10. Records Retention

- 3.10.1. All Personal Data shall be stored securely by the Provider until such time as the outputs have been prepared for the Customer after which the Personal Data shall be destroyed securely in line with the data disposal policy standards for “OFFICIAL-SENSITIVE” information of the United Kingdom Government. The Provider must delete all records within 6 months of processing the data. However, read-through indexes may be retained until such point as the Customer instructs the Provider to delete them.

3.11. Business Continuity

- 3.11.1. To allow for business continuity, the Provider may, where appropriate and as agreed with the Customer, produce an index to allow for read-

through to support efficient and effective linkage of future SILC projects.

3.12. *Audit Rights*

- 3.12.1. The Provider agrees that the Customer can, upon giving reasonable notice and within normal business hours, carry out compliance and information security audits and checks to ensure adherence to the terms of this MoU.

3.13. *Quality Assurance*

- 3.13.1. The Provider must take all reasonable measures to ensure the quality assurance of the processed data.

3.14. *Constraints*

- 3.14.1. The Provider must not, during and after the term of this Agreement:
 - a) use any Contract Personal Data other than as directed by the Customer;
 - b) use any Contract Personal Data for its own direct or indirect benefit, or the direct or indirect benefit of any third party, except that the Provider may use Contract Personal Data to the extent necessary to perform its duties and obligations, or to enforce its rights, under the Contract;
 - c) allow Contract Personal Data to be accessed by, or sent to, parties outside the EEA (unless expressly required or permitted to do so by the Customer);
 - d) seek to gain commercial advantage from its access to Contract Personal Data;
 - e) disclose any Contract Personal Data to third parties, other than as required by the terms of this Agreement or as required by a court or other competent authority in which case the Customer and the Provider should discuss together the appropriate response to any and all such requests.

4. Termination and Exit

4.1. *Termination Rights*

- 4.1.1. Either party shall be entitled to terminate this Agreement by giving not less than **6 months** written notice.

4.2. **Termination Process and Responsibilities**

- 4.2.1. Within 30 calendar days following termination of this Agreement by either party, the Provider shall, at the direction of the Customer:-
- a) comply with any other agreement made between the parties concerning the return or destruction of Personal Data (Section 3.10);
 - b) return all Contract Personal Data passed to the Provider by the Customer; or
 - c) on receipt of instructions from the Customer, destroy all Contract Personal Data in its possession or control unless prohibited from doing so by any applicable law, and confirm in writing to the Customer that it has done so.

5. Charging and Payments

- 5.1. “The Customer” has already provided payment for up to 30 projects which it has approved to be processed by “the Provider”. This MoU has been set-up to support delivery of these projects, with “the Provider” undertaking the processing of data as required to support the linkage in-line with the Data Linkage Framework, the Guiding Principles for Data Linkage and the broader Scottish Informatics and Linkage Collaboration (SILC) agenda.

6. Signatories

Service Customer:

Organisation: Scottish Government, Division Name

Name: Name Of Information Asset Owner – Deputy Director/Head of Division

Position:

Signature: _____ Date: __ / __ / __

Service Provider:

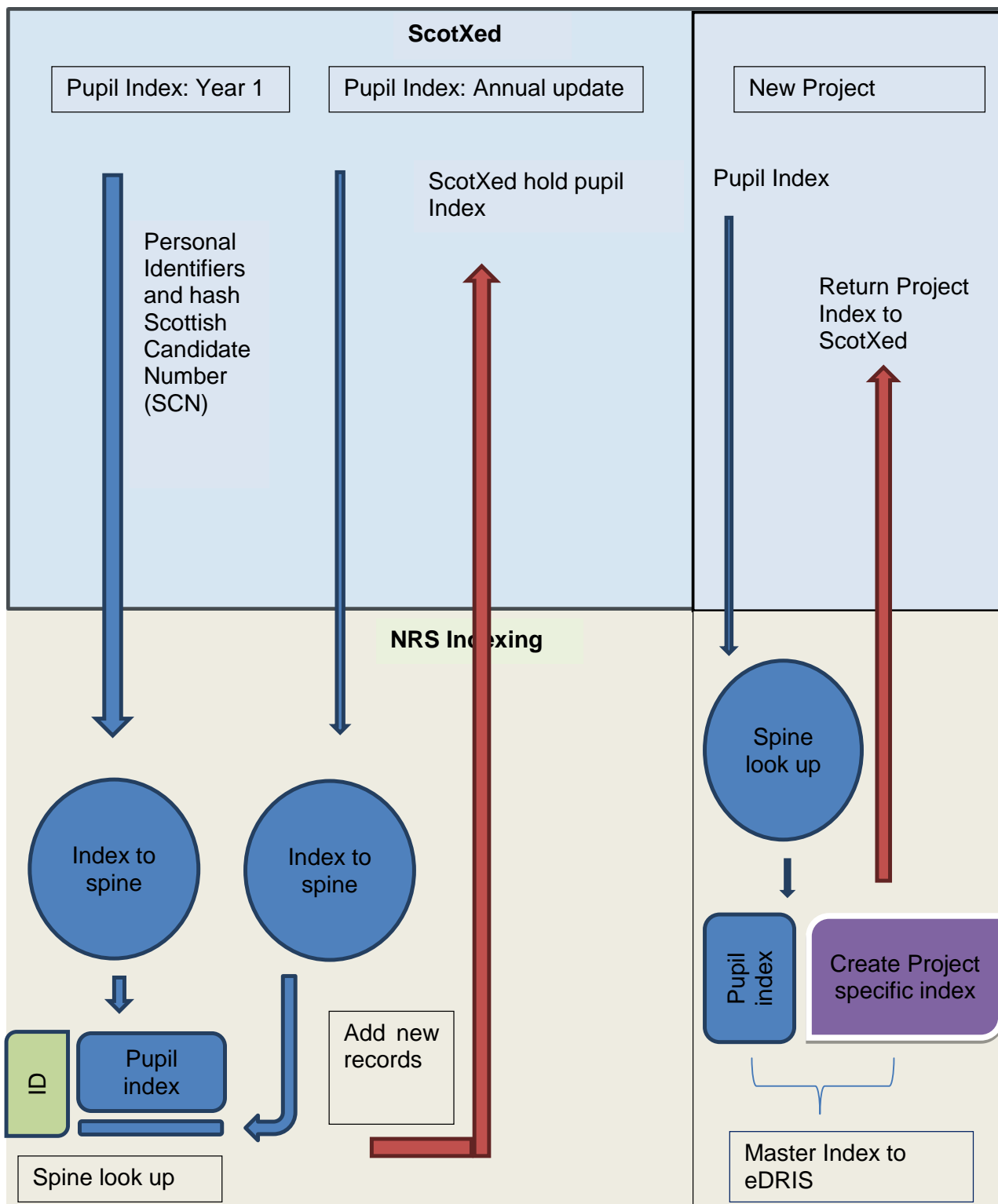
Organisation: National Records of Scotland (NRS)

Name:

Position:

Signature: _____ Date: __ / __ / __

Education Read Through diagram



8.3

UK Statistics Authority
National Statistician's Data Ethics Advisory Committee
NSDEC(16)08
Paediatric National Survey of Bereaved People (VOICES)

This proposal is in draft form and will be published in due course

Any other business

10