



**UK Statistics
Authority**

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

Minute, Agenda and Papers

Tuesday 25 April 2017

10:30 – 11:45

Board Room, UK Statistics Authority
London

UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE

Agenda

Tuesday 25 April
Board Room, Drummond Gate London
10:30am – 1:45pm

Chair: Mr Ian Cope

Apologies: Annie Hitchman, Keith Dugmore, Neil McIvor

(10:30am to 12:00pm)

1 10:30am	Minute and matters arising from the previous meeting (Page)	Mr Ian Cope
2 10:40am	Chair's report (Page)	Oral Report Mr Ian Cope
3 11:00am	Approving applications by precedent	Oral Presentation Mr Ian Cope
4 11:15am	Review of Terms of Reference	NSDEC(17)07 Mr Ian Cope
5 11:30pm	ADRN: Veterans in Great Britain: Linking administrative data from the Veterans Leavers Database with Census 2011	NSDEC(17)08 Mr Jon Wroth-Smith Rob Powell
6 11:45pm	LFS: Project to assess the representativeness of the Labour Force Survey Phase I	NSDEC(17)09 Mr Ed Dunn

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

(12:30pm to 1:45pm)

7 12:30 pm	MRP: The societal impacts of water problems (such as flood and supply interruption) on health and wellbeing	NSDEC(17)10 Mr Tony Chapple
8 12:40 pm	MRP: Study of Small and Medium-sized Enterprises (SMEs) in the UK as part of the European Commission's SME Performance Review 2016/17	NSDEC(17)11 Mr Tony Chapple
9 12:55pm	MRP: Impact evaluation (Phase I and II) of the Innovate UK Technology and Innovation Centre network (Catapults)	NSDEC(17)12 Mr Tony Chapple
10 1:10pm	Department for Education's approach to data transparency and privacy	Oral Presentation Mr Iain Bell
11 1:40pm	Any other business	

Next meeting: Tuesday 4 July 2017

Location: One Drummond Gate, London

National Statistician's Data Ethics Advisory Committee

Minute

Tuesday, 24 January 2017
Board Room, Drummond Gate, London

Present

Members

Mr Ian Cope (Chair)
Ms Vanessa Cuthill
Mr Colin Godbold
Ms Annie Hitchman
Mr Neil McIvor
Ms Isabel Nisbet
Ms Marion Oswald
Mr Osama Rahman

UK Statistics Authority

Mr Adil Deedat
Dr Simon Whitworth

Office for National Statistics

Mr Peter Stokes (for item 3 to item 6)
Mr Matthew Greenaway (for item 7)
Ms Lan Benedikt (for item 8)
Mr David Johnson (for item 8)

Scottish Government

Dr Andrew Waugh (for item 9)

Apologies:

Mr Robert Bumpstead
Mr Keith Dugmore
Dr Dean Machin
Professor Martin Severs

1. Minutes and matters arising from the previous meeting

- 1.1 The Chair welcomed members to the seventh meeting of the National Statistician's Data Ethics Advisory Committee (NSDEC). The Chair informed the committee that Hetan Shah has stepped down as a member of NSDEC ahead of his appointment ending in July 2017.
- 1.2 Members were informed that the minute of the sixth meeting had been agreed by correspondence. The minute, agenda and papers from the last meeting are now published on the [UK Statistics Authority website](#).
- 1.3 The Chair updated the meeting with progress on actions from previous meetings. Most actions were complete or in progress and would soon be complete.

2. Chair's report

- 2.1 The Chair informed the meeting of ongoing engagement to raise the profile of NSDEC and its work to date. Members heard that the Chair had blogged on the ONS internal portal, informing ONS staff how ethical review could benefit business areas, the committee's work to date and feedback he had received from business areas who had previously engaged with the committee.
- 2.2 The Chair provided members with an update on projects previously considered by NSDEC. The meeting heard that the Welsh Government proposal to analyse de-identified linked Census and National Survey for Wales data, had had all minor revisions implemented. The paper has now been published on the UK Statistics Authority Website. ONS and the Secure Anonymised Information Linkage Databank (SAIL) (both Trusted Third Parties within the Administrative Data Research Network (ADRN)) are now working together to link and de-identify the data before it is made available to Welsh Government researchers within the secure environment.
- 2.3 The second proposal from Welsh Government, which aimed to understand the outflow of Welsh speakers from Wales to England by linking Welsh Census Records to the 2011 Census, received major revisions. Members heard that all revisions required from the researcher had been implemented; however ONS are still waiting for consent from the Secretary of State for Health to onwardly disclose de-identified linked Patient Register data to the Administrative Data Research Network. Should the Secretary of State for Health give consent then the application will be reconsidered by NSDEC as it is currently under major revisions.
- 2.4 The Census Transformation Programme (CTP) presented to the last meeting on their work to date producing income estimates. The meeting heard that CTP would be presenting an application to the next meeting detailing their existing work and their future plans to produce household income estimates with greater granularity.
- 2.5 The proposal from the British Broadcasting Corporation (BBC) to link Census and TV Licensing data will no longer take place. Following ethical approval, the BBC decided not to pursue this research.
- 2.6 All recommendations relating to the Ethnicity from names project are almost complete. At the April meeting members will receive a paper from ONS which details how each of the major revisions have been addressed.
- 2.7 The Chair updated the committee on Data Access Legislation which forms part of the Digital Economy Bill. The meeting heard that the Bill will soon be considered in the Lord's Committee Stage.
- 2.8 The Chair concluded his report by providing a comparison of the Government Digital Service's (GDS) ethical framework, principles and scope and processes with those of NSDEC. Members heard that the Chair will continue engaging with those developing a second iteration of the GDS framework as part of his role on the Data Leader's Network.

3. MRP: Formalising PriceWaterhouseCoopers (PWC) request to use ONS data [NSDEC(17)01]

- 3.1 The Chair invited Mr Peter Stokes to the meeting. This Approved Researcher (AR) proposal had previously been considered by correspondence. Mr Stokes thanked the committee for expediting this application. This was the first AR application from a commercial organisation following the updating of the AR criteria and guidance.

- 3.2 The meeting heard that there had been follow up meetings with PriceWaterhouseCoopers (PWC) and the commissioning body, London First, who are due to publish the report in February.
- 3.3 In light of this application, members suggested that the Approved Research scheme publishes guidance for approved researchers relating to how analyses are communicated as well as publication timetables. Consideration should also be given to how any changes in publication timetables are managed. This should include contingencies that can be quickly put in place when ONS are not notified in advance of changes to publication timelines.
- 3.4 It was suggested that the Microdata Release Panel (MRP) which governs access to de-identified ONS data through the secure Virtual Microdata Laboratory (VML) consider whether the Approved Researcher process should take in to account whether the body from which a commercial researcher originates should also be assessed to determine whether they are a fit and proper organisation.
- 3.5 It was suggested that the MRP also consider asking Approved Researchers to include their quality assurance processes in their proposals.

Action: Mr Peter Stokes to consider including additional guidance for approved researchers as detailed and provide an update at the next meeting.

4. Applications: Policies and process for expedited review [NSDEC(17)02]

- 4.1 The Chair introduced the item. Following expedited review of an Approved Researcher proposal referred by the Microdata Release Panel, the Chair informed the meeting that consideration of some proposals via correspondence may be necessary in order to ensure equality in timeliness of access to ONS data.
- 4.2 Members agreed that option one, where all members are sent the application for review would be most appropriate and worked well in the case of reviewing [NSDEC(17)01].
- 4.3 Members agreed that, in order for any expedited proposal to be formally approved, a sufficient number of members would need to respond in accordance with a quorate meeting.
- 4.4 It was agreed that where expedited review is sought and NSDEC approve the proposal, the application and summary of member's comments will be published as part of NSDEC's next meeting.

5. MRP: Frontier Economics use of ONS data [NSDEC(17)03]

- 5.1 Mr Peter Stokes provided an overview of the Approved Researcher application. The meeting heard that Frontier Economics, on behalf of Tech City UK, would be replicating an analysis undertaken last year which was published within the Tech Nation UK report.
- 5.2 Members suggested that the public good be clarified. They also sought clarity around the methods employed and suggested consideration of other data sources and methods which could provide a more complete picture of the UK Tech Sector. This could include use of Company's House data or employing methods such as text mining.
- 5.3 NSDEC agreed that the proposal could proceed subject to clarifying the public benefit and considering the use of alternative data sources and methods which may provide a more complete picture of the tech sector.

Action: Mr Pete Stokes to amend the application to clarify public benefit and to discuss with researchers the use of alternative data sources and or methods which may provide a more complete picture of the tech sector.

6. MRP: Road safety analysis Ltd use of ONS data [NSDEC(17)04]

- 6.1 Mr Peter Stokes introduced the Approved Researcher proposal which would see Road Safety Analysis Ltd (RSA) use ONS mortality data to produce an analysis for Highways England on road related suicides.
- 6.2 Members suggested that the analysis could also benefit from investigating attempted road suicides using other data sources.
- 6.3 NSDEC identified that there was a small risk of harm to researchers using the data who may come across an individual they have known in the data set but were unaware of their cause of death. In order to minimise risk of harm to researchers it was suggested that researchers are made aware of this when they are analysing the data.
- 6.4 Overall members agreed that the proposal demonstrated a clear public benefit. However members were unclear as to why home postcode of the deceased at the time of death was required for this analysis.
- 6.5 NSDEC agreed that the proposal could proceed subject to reviewing the need for home postcode information.

Action: Mr Stokes to clarify the need for home postcode information and share reasoning with the committee via correspondence.

7. Web-scraping guidance [NSDEC(17)05]

- 7.1 The Chair welcomed Mr Matthew Greenaway to the meeting. Mr Greenaway provided an overview of the proposed guidance which had been compiled to formalise good practices in web-scraping for use by ONS staff in the production of statistics and research which serve the public good.
- 7.2 The meeting heard that in order for ONS to undertake web-scraping on a large scale, i.e. scraping small amounts of data from a large number of websites, there were challenges in checking website terms and conditions and gaining consent from website owners.
- 7.3 Members were generally supportive of the guidance, but suggested further work be undertaken on the guidance.
- 7.4 This included considering the use of text mining and use of machine learning techniques in order to automatically check terms and conditions of websites to understand whether they can be legally scraped.
- 7.5 It was suggested that references to personal data needed to be defined with an emphasis that any proposal to scrape personal data from a website be considered by NSDEC before scraping is undertaken.
- 7.6 It was suggested that ONS may want to undertake engagement with relevant businesses whose websites they are most interested in scraping. The guidance should also provide greater clarity as to what is meant by large amounts of data.
- 7.7 It was suggested that greater clarity be sought as to the legal implications of breaching terms and conditions of website owners. Members referenced a case considered in the EU Court of justice. It was suggested that the Big Data team consider the implications of this ruling and related copyright and data protection in redrafting the guidance.
- 7.8 Further consideration should also be given to the types of websites being scraped to ensure they themselves are not unethical and do not pose a risk to the ONS

Action: Secretariat to work with Mr Greenaway to reflect NSDEC's comments in to the application and to bring back to the committee at a future meeting for further consideration.

Action: The Chair in his role as an ONS executive to offer support to establish a task and finish group to further refine the draft guidance.

8. Big Data to Tourism Statistics

8.1 The Chair welcomed Ms Lan Benedikt and Mr David Johnson from the ONS Data Science Campus to the meeting. The Data Science Campus is set to act as a hub for the whole of Government to gain practical advantage from the wider investment in data science research and help cement the UK's reputation as an international leader in data science.

8.2 Ms Benedikt set out existing practices for producing tourism and migration statistics using the International Passenger Survey (IPS). The meeting heard of limitations to the IPS, such as sampling error. There are a limited number of data sources currently available to ONS to supplement estimates produced using the IPS.

8.3 In order to address this, the Data Science Campus is working with other partners in government, academia and from other National Statistics Institutes to consider how a range of open data sources could be used to the produce tourism and migration statistics.

8.4 The meeting heard of initial feasibility work using geo-located Flickr data. Here, through the Flickr API, user data is downloaded. All identifiers which could be used to identify an individual are removed and the data is aggregated to show where groups of tourists are localised around particular landmarks. Ms Benedikt stressed to members that photographs uploaded by Flickr users were not downloaded.

8.5 Members were divided in their views on the use of geo-located social media data. There were concerns that whilst individuals may have agreed to sharing their location they may not have been informed in providing consent for reuse of social media user's data. Other members also expressed concerns around minors breaching the terms and conditions of use and registering to have accounts.

8.6 Ms Benedikt informed members that research using social media data was purely exploratory at this stage and would contribute to further developing methods and understanding how big data could be used in the production Official Statistics.

8.7 Members heard that the Data Science Campus would be working with the ONS Big Data team to develop guidelines for how social media should be used by ONS for research and statistics which serve the public good.

8.8 NSDEC members noted the difference between initial feasibility research and the use of social media data in the production of official statistics. It was suggested that this research could proceed for the time being; however the Data Science Campus should consider providing a paper or application to NSDEC addressing each of NSDEC's ethical principles with regards to the use of social media data.

Action: the Data Science Campus and Big Data team to develop guidance for the use of social media data for research and statistics which serve the public good.

Action: the Data Science Campus to produce a paper or application detailing their plans for use of social media data to produce experimental statistics.

9. National Records of Scotland: using public and commercial sector administrative data in the production of population statistics [NSDEC(17)06]

- | |
|-----|
| 1. |
| 2. |
| 3. |
| 4. |
| 5. |
| 6. |
| 7. |
| 8. |
| 9. |
| 10 |
| 11. |
- 9.1 The Chair welcomed Dr Andrew Waugh from the National Records of Scotland to the meeting.
- 9.2 Dr Waugh outlined the NRS's ambition to use data from commercial companies who were listed on Crown Commercial Services; this could include the use of data held by credit reference agencies for the production of population statistics.
- 9.3 Dr Waugh explained that as government administrative data are collected for administrative purposes there were challenges using it in the production of statistics. For instance 30-59 year old males are often over represented on government sources. This may be the result of these groups either not interacting with GP services or who are no longer resident and not notified DWP of a change in their circumstances. Through considering credit reference agency data it may be possible to determine whether or not individuals are still active in the country.
- 9.4 NSDEC saw the benefits such data could bring to the production of administrative data based population estimates; however there were reservations regarding the ethics of commercial companies profiting from the sale of the public's data which had been provided for another specific purpose (credit checking). Consideration was given to whether the consent would only relate to that specific use.
- 9.5 Members suggested further engagement with DWP and HMRC to explore ways in which their data could be used, in a similar way to how credit reference agency data use was proposed. It was also suggested that NRS work closely with ONS and NISRA to ensure a joined up approach, especially around approaches to commercial data sources.
- 9.6 Members agreed that whilst the proposal had clear public benefits more work should be undertaken to explore existing data sources in government. Discussions with other UK statistical institutions should also be undertaken before considering procurement of commercial datasets. NSDEC therefore recommend major revisions to this proposal whilst these conversations take place.
- Action: Dr Andrew Waugh to discuss potential of existing government sources and procurement of commercial data sources with ONS**
10. **Any other business**
- 10.1 There was no other business.

Chair's report

Mr Ian Cope

1.											
	2.										
		3.									
			4.								
				5.							
					6.						
						7.					
							8.				
								9.			
									10		
										11.	

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

Applications: Policies and process for approval by precedent

Oral presentation

Ian Cope

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.
----	----	----	----	----	----	----	----	----	----	-----

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(17)07

Updated terms of reference of the National Statistician's Data Ethics Advisory Committee

Purpose

1. This paper presents some updates to the National Statistician's Data Ethics Advisory Committee's (NSDEC) terms of reference. These updates are presented in red in Annex A.

Recommendations

2. Members of NSDEC are invited to consider and agree the proposed updates to the terms of reference.

Background

3. At the meeting of the Board of the UK Statistics Authority on 6 November 2014, the Board agreed that NSDEC should be established. The Board noted that this committee should be advisory to the National Statistician, contain independent members and be guided by a set of principles established and owned by the Board.
4. The terms of references of NSDEC were discussed and agreed at the meeting of the committee on 14 October 2015.

Updates to the Terms of Reference

Approved Researcher Scheme

5. NSDEC have agreed that NSDEC would provide ethical approval for proposals to access ONS data from the commercial sector via the Approved Researcher Scheme. These requests are referred to NSDEC for ethical consideration by the Microdata Release Panel (MRP) which governs access to the Virtual Microdata Laboratory (VML).
6. Existing provisions within the Approved Researcher scheme allow academics to access ONS data following approval of their project by the Microdata Release Panel (ethical review is not mandatory for academic projects although often a review has been undertaken by their host institution).
7. The additional requirement for commercial organisations to gain ethical review from NSDEC could mean that commercial access to data may be less timely than requests originating from academia. Therefore, it was agreed in the NSDEC meeting on 24 January that NSDEC will consider some of these proposals from the commercial sector by correspondence.
8. In these cases it was agreed that there must be a clear reason for requiring an expedited ethical review by correspondence. For example this may be:
 - i. where the benefits of the proposal to use ONS data through the approved researcher scheme cannot be realised by waiting for the next NSDEC meeting; and
 - ii. where waiting for the next meeting means timeliness of access to ONS data through the approved research process, is hindered.

Administrative Data Research Network (ADRN)

9. At the request of the chair of the ADRN Board, NSDEC provides ethical consideration for government and third sector researchers wishing to access data via the ADRN. At the NSDEC meeting on 11 October, members reviewed its first two ADRN projects from Welsh Government.
10. Before being considered by NSDEC, ADRN projects require approval from the ADRN Approvals Panel which meets monthly. In order to ensure ADRN projects can proceed as quickly as possible it was agreed in the NSDEC meeting on 24 January that there will be circumstances where ADRN projects are considered between meetings.
11. Expedited ethical review by correspondence of ADRN projects will only be considered in the following cases:
 - i. where ADRN projects have been approved by the approvals panel (which meets monthly) shortly after an NSDEC meeting; or
 - ii. where an ADRN project has been approved by the approvals panel but received major revisions at a previous NSDEC meeting.
12. The Secretariat has updated the terms of reference and an updated version is presented at **Annex A**. The terms of reference will be reviewed as part of NSDEC's self assessment in July.

Simon Whitworth, NSDEC Secretariat, Central Policy Secretariat, UK Statistics Authority, 3 April 2017

List of Annexes

Annex A: Updated NSDEC Terms of Reference

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

Annex A: Updated NSDEC Terms of Reference

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

Terms of reference

Introduction

1. At the UK Statistics Authority Board meeting on 6 November 2014, the Authority Board agreed to establish the National Statistician's Data Ethics Advisory Committee (NSDEC), which would be advisory to the National Statistician.

Role and responsibilities

2. The role of the National Statistician's Data Ethics Advisory Committee (NSDEC) is to advise the National Statistician on the ethical considerations concerning the accessing, sharing and use of data.
3. Specific responsibilities of the NSDEC include to:
 - i. provide ethical consideration of proposals to access, share and use data;
 - ii. advise on individual policies and research projects against NSDEC's ethical principles;
 - iii. develop a consistent ethical framework for relevant projects related to official statistics;
 - iv. promote transparency around data shares;
 - v. provide ethical approval for some government, and third sector researchers who wish to use ONS data for research and statistics that serves the public good;
 - vi. provide ethical approval for proposals to access ONS data from the commercial sector via the Approved Researcher Scheme. These requests are referred to NSDEC for ethical consideration by the Microdata Release Panel (MRP) which governs access to the Virtual Microdata Laboratory (VML); and
 - vii. provide ethical advice on other data issues within the National Statistician's remit, which includes his role as head of the Government Statistical Service.
4. At the invitation of the Chair of the Administrative Data Research Network (ADRN), provide ethical consideration for government and third sector researchers wishing to access data via the ADRN. These requests are referred to NSDEC by the ADRN Approvals Panel.
5. In reviewing proposals NSDEC will provide advice by consensus.
6. To facilitate timely access to data NSDEC can consider proposals referred to it by the MRP or the ADRN Approvals Panel by correspondence.

Meetings

7. NSDEC will meet at least four times a year. The Chair of NSDEC may convene additional meetings as deemed necessary.
8. The Chair may invite other relevant experts as appropriate to advise the committee.

Reporting

9. NSDEC will provide reports following each of its meetings to the National Statistician.
10. At least seven days prior to its meetings, NSDEC will receive reports on:
 - i. proposed new data shares;
 - ii. relevant projects and programmes and other information as it requires; and
 - iii. reports from any sub-committees.
11. NSDEC will operate transparently. Meeting agendas, papers and minutes will be made publicly available on the UK Statistics Authority website.
12. Where expedited review is sought and NSDEC approve the proposal by correspondence, the application and summary of members' comments will be presented at NSDEC's next meeting and published as part of the meeting papers.

Membership and role of members

13. The members of NSDEC will include:
 - i. a representative of the National Statistician (Chair);
 - ii. at least five independent external members, including lay members; and
 - iii. no more than four members from government.

Quorum

14. NSDEC meetings will be considered quorate when four or more members are present including at least two independent external members as well as the Chair or the Chair's delegated nominee.
15. Substitutes will not be permitted to attend meetings, unless with the invitation of the Chair.
16. In order for any expedited proposal to be formally approved by correspondence a sufficient number of members would need to respond in accordance with a quorate meeting.

Secretariat

17. Secretariat for the NSDEC will be provided by the staff from the Central Policy Secretariat within the UK Statistics Authority.

Review

18. NSDEC will review the effectiveness of its meetings and its terms of reference annually.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(17)08

ADRN: Veterans in Great Britain: Linking administrative data from the Veterans Leavers Database with Census 2011

Purpose

1. This paper presents a proposal to link information from the MoD Veteran Leavers Database with the 2011 Census in England, Wales, Scotland and Northern Ireland. The proposed linkage will be used to establish an evidence base at a lower super output area level to inform the public and voluntary sector to improve existing services and target new services to veterans and their households.

Recommendations

2. Members of NSDEC are invited to consider the 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 the proposal be resubmitted to a future meeting once implemented; or
 - iv. reject the proposal advising it be stopped from proceeding.

Background

3. The Armed Forces Covenant sets out the relationship between the nation, the government and the armed forces. It is a promise from the nation that those who serve or have served, and their families, are treated fairly. The Ministry of Defence (MoD) are working with businesses, local authorities, charities and community organisations to support the forces through services, policy and projects. The covenant recognises that the whole nation has a moral obligation to members of the armed forces and their households.
4. There is interest from several public and voluntary sector organisations to develop a better understanding of veterans in the UK. There is also interest in veteran households who are often overlooked by researchers and policymakers which can result in social isolation, disrupted health and social care treatment, disrupted education for children and difficulties for services sharing data for safeguarding activities.
5. The Ministry of Defence publishes statistics on the number of UK veterans in receipt of compensation or pensions by location. However estimates from the ONS Annual Population Survey indicate a significantly higher number of veterans living in the UK than the MoDs statistics.
6. Following a public consultation on information needs for the 2021 Census, which generated nearly 1100 responses, the Office for National Statistics (ONS) published their current view (May 2016) on the topics proposed for inclusion in the 2021 Census. The ONS view about information on the armed forces community was that there was a clear need for this information but that the need may be better met through alternative sources, such as linking administrative data or the Annual Population Survey.

7. Since then, ONS has received more feedback on the ONS decision not to record veteran service personnel in the 2021 Census. This has led ONS and the Ministry of Defence (MoD) to establish a working group to better understand the need and the best way to meet that need. The group is meeting regularly to develop and progress 3 potential themes to best meet the information need:
 - I. Theme 1 is focusing on better articulating the user requirement for information on armed forces leavers. Input has been sought from across government, local authorities, health care bodies and the charitable sector.
 - II. Theme 2 is focusing on making better use of existing data to provide information on the needs of armed forces leavers. ONS and MoD are developing a proposal and working through the necessary approvals as a test case for data linkage. This includes the challenges of data release, data security, data matching and analysis. The existing data work is looking at several options. These include anonymously linking MoD administrative data sources to the 2011 Census and further analysis of information collected about armed forces leavers from the Annual Population Survey.
 - III. Theme 3 is focusing on census question development. This includes developing and testing a question on armed forces leavers that:
 - a. would work in the context of the census;
 - b. is readily understood by respondents; and
 - c. would produce reliable information.
8. This application is relevant for theme 2 and will explore the linkage of the MoD Veteran Leavers Database with the 2011 Census.
9. The successful delivery of this project will enable public and voluntary organisations to deliver improved services as described in the Armed Forces Covenant. The delivery of these services advocates a clear public benefit towards a specific population group (veterans and veteran households) and depends on a comprehensive evidence base which will be provided by this project.

Petros Saravakos, NSDEC Secretariat, Central Policy Secretariat, UK Statistics Authority, 10 April 2017

List of Annexes

Annex A: Application: Veterans in Great Britain: Linking administrative data from the Veterans Leavers Database with Census 2011

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

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

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

Section A
Application Details

A1	Responsible Owner
<p>Full Name: [REDACTED]</p> <p>Address: [REDACTED]</p>	<p>Position: [REDACTED]</p> <p>Email: [REDACTED]</p> <p>Telephone: [REDACTED]</p> <p>Organisation: [REDACTED]</p>
<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:</p> <p>Date: [REDACTED]</p>	

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

A2 Applicant Details (if applicant is not the responsible owner)

Full Name:	Position:
Address:	Email:
	Telephone:
	Organisation:

A3 Project Information

Project Title: Veterans in Great Britain: Linking administrative data from the Veterans Leavers Database with Census 2011

Start Date: 26/04/2017

End Date: 01/012018

Project Sponsor (select all that apply)

- ☒ ONS
- ☒ ADRN
- ☐ GSS
- ☐ Collaboration
- ☒ Other (Please specify).....Ministry of Defence (MoD)

A4 Collaboration and Sponsors

List of Collaborators/Sponsors	Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)
Office for National Statistics	ONS, MOD, NRS and NISRA are data suppliers. ONS will provide the Trusted Third Party data linkage service on behalf of the ADRN.
Ministry of Defence (MoD)	
Administrative Data Research Network (ADRN)	
National Records Scotland (NRS)	
Northern Ireland Statistics and Research Agency	

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 (please specify).....

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)

The Armed Forces Covenant (2011) was introduced to recognise the contributions of the armed forces community by redressing any challenges they might face. Although the covenant does not establish a legal framework to address these challenges its key principles have been enshrined in law within the Armed Forces Act 2011. Therefore there is a need for organisations to comply and demonstrate delivery of the improvements outlined in the covenant.

Specifically the covenant sets out the relationship between the nation, the government and the armed forces. It recognises that the whole nation has a moral obligation to members of the armed forces and their households, and it establishes how they should expect to be treated. This is reflected in the covenant's two principles:

- i. the armed forces community should not face disadvantage compared to other citizens in the provision of public and commercial services; and
- ii. particular consideration is appropriate in some cases, especially for those who have given most such as the injured and the bereaved,

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

The covenant¹ focuses on helping members of the armed forces community have the same access to government and commercial services and products as any other citizen. The scope of the covenant includes the following areas:

- education and family well-being
- having a home
- starting a new career
- access to healthcare
- financial assistance
- discounted services

The covenant is fulfilled by different groups including:

- Central government
- Single services (Royal Navy, British Army, Royal Air Force)
- Businesses of all sizes
- Local government
- Charities
- Communities
- Cadet forces and their adult volunteers

To deliver this role the AFC includes several covenants:

- **Covenant for businesses:** Businesses and charitable organisations who wish to demonstrate their support for the armed forces community can sign the covenant. In doing so they make a range of written and publicised promises to set out their support
- **Covenant for communities:** For communities, the Armed Forces Covenant is a promise of mutual support between a civilian community and its local armed forces community. A Covenant Fund is available for community projects that offer small grants up to £20,000 and large grants up to £500,000.
- **Covenant for serving personnel and families:** There are several initiatives as part of the Armed Forces Covenant designed to support serving personnel and their families. There is also information about understanding reservists' employment rights.
- **Covenant for veterans:** practical support and companionship to improve older veterans' wellbeing and services to assist individuals in their needs.
- **Covenant for policy makers:** provides information to policymakers who wish to support the armed forces community through the AFC

In light of this covenant, there has been a broader demand for information on veterans so that services can be targeted to meet their needs. This includes public services as well as private (non-commercial) organisations (e.g. charity organisations).

Currently the MoD publishes numbers of UK veterans by location for those in receipt of compensation or pensions (which is approximately 374,000). However, estimates from the ONS Annual Population Survey (APS) indicate that there are 2.6 million veterans living in the UK. Therefore, the numbers of veterans by the MoD methods/sources are not sufficient to provide an evidence base of the geographic location or demographic characteristics of veterans required to allow organisations to comply with the Armed Forces Covenant. In addition, as a result of the relatively small sample sizes from the APS it is not possible to

¹ <https://www.armedforcescovenant.gov.uk/about/>

<https://www.gov.uk/government/policies/armed-forces-covenant>

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

provide cross tabs with other characteristics at a local level (e.g. LA level or lower).

In developing our understanding of veteran's information needs it has also become clear that we need to identify household members as this group are also covered by the Community Covenants but are often overlooked which can result in social isolation, disrupted health and social care treatment, disrupted education for children, and difficulty for services in sharing data for safeguarding activities. Please note, through using Census data it will only be possible to identify dependents/families living in the same household.

In response to this growing demand, the Office of National Statistics (ONS), National Records Scotland (NRS), Northern Ireland Statistics and Research Agency (NISRA) and the Ministry of Defence (MoD) have since set up a working group to examine options for delivering data to assist organisations in demonstrating the delivery of the Armed Forces Covenant. One of the focuses of this working group is to explore whether the use of administrative data will provide the level of geographic detail required (at lower or middle layer super output areas). This project has been approved by the Administrative Data Network Approvals Panel.

There are two strands to this project:

1. explore the feasibility of linking the Veteran Leavers Database (VLD) held by the Ministry of Defence (MoD) with the 2011 Census data held by the Office for National Statistics (ONS), National Records of Scotland (NRS), and the Northern Ireland Statistics and Research Agency (NISRA). This will benefit future research projects by providing the methods and tools to effectively and responsibly link these datasets.
2. if the linkage is successful it will provide an evidence base which will enable public services to:
 - i. understand the demographic characteristics of veterans and veteran;
 - ii. households living in UK and allocate appropriate resources and improved services;
 - iii. comprehend the geographical distribution of veterans and veteran households living in UK to develop local services and deliver targeted initiatives;
 - iv. establish whether VLD data quality supports future administrative data linkage projects; and
 - v. support new publications on information relating to veterans.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

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>			Veterans Leavers Database, Ministry of Defence	
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>			England & Wales Census 2011, ONS Scotland Census 2011, NRS Northern Ireland Census 2011, NISRA	
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?
(max 500 words)

ONS will provide secure data linkage and a secure environment in which approved researchers can access only the linked data relevant to their approved projects.

In particular, full account has been taken of the risks associated with:

- transfer of administrative data from data owner to ONS;
- securing access to the data to be linked;
- the management of both the identifying data and attributes contained within the administrative data sources;
- the linkage process;
- the identification of population subgroups within the linked data;
- transfer of the de-identified linked file to a secure environment; and
- secure access by researchers at the secure environment.

This is being done via the following controls:

- **Physical** – all linkage takes place behind high-security doors in a secure physical environment.
- **Procedural** – all data acquisition, import and export processes are subject to strict procedural controls.
- **Personnel** – only authorised employees holding Security Check clearance are permitted to enter the environment, and all access is recorded, monitored and audited by ONS Security Managers on a regular basis, through regular review of technical, procedural and CCTV records.
- **Technical** – the linkage facility is fully isolated from all other systems and networks. Within the environment, technical safeguards exist to ensure only authorised work can take place, and “unusual” activity is detected, assessed and acted upon. Electronic devices, (including mobile phones), software or connections are not permitted in the environment under any circumstances, and protective measures are in place to enforce this policy.

The linked attribute dataset will be protected at all times whilst in transit from the linkage facility to the VML. Access to the data via the VML is only possible through controlled rooms, located in ONS and other Government buildings. Before access is made available at any given location, it has to be approved by ONS.

Researchers will take no data into the room, and will take out no data. Results from the research will only be released by ONS staff once they have confirmed that they contain no risk of identifying an individual, household or business.

Rather than operating a traditional 'rules based' approach to output checking, the VML operates a 'principles based approach'. This forms part of the “5 safes approach” to data access adopted by the VML. <https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/>

To summarise, research outputs created within the VML are checked against threshold rules. When an output fails the threshold rule (for example a cell count less than 10) a discussion with the researcher takes place, giving the researcher an opportunity to demonstrate the output is not disclosive and therefore safe. By applying a principles based approach, rather than a black and white rule, context can be taken into consideration,

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

however ultimately, it is the VML statistical support team that make the final decision.

To limit the risk of re-identification we have introduced the following safeguards:

- we are compliant with government guidelines and best practice on assessing the risk of re-identification. The Government Communications Headquarters are content with the approach and controls in place;
- any future re-use of data will only take place following a completed risk assessment as described above; and
- any outputs will be subject to statistical disclosure controls and no aggregate outputs below the minimum cell count of 10 will be published;

ONS will retain data for 12 months after delivering the dataset in case there is a requirement to carry out follow-up linkage work. The ADRN is currently reviewing a policy of retaining data, and the ADRN will report back to NSDEC once this policy has been agreed.

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

The Ministry of Defence (MoD), together with other government departments, the devolved administrations, partner charities and voluntary organisations, have been working to fulfil the series of commitments made to the armed forces community.

This research aims to examine veterans of the UK Armed Forces eligible to remain in the UK. The Defence Statistics has a Veteran Leavers Database (VLD), a single source of electronic information for Service personnel that have left the UK Armed Forces, irrespective of regular/reserve status and length of Service. This database contains records for approximately 1.8 million Service leavers, sourced from a variety of legacy and current administrative systems held within the MoD. The information in the VLD reflects all veterans who have left the UK Armed Forces since circa 1970. Data are downloaded from the Joint Personnel Administration system (JPA) and legacy systems (including the Armed Forces Pay and Pensions Administration, AFPAA).

Census data collected in 2011 for all responding households in England and Wales; Scotland; and Northern Ireland (approximately 94% of all households).

The Veteran Leavers Database (VLD) and Census will be linked on the basis of the following variables:

- surname
- previous surname
- forename(s) or initials
- date of birth
- gender

The following information will be retrieved from the 2011 Census data:

- Structure of household: information about the household, for example the type of residence and accommodation, tenure, sleeping rough, communal establishment, household deprivation score, living arrangements and household composition. This information will be used to determine whether veterans have the same access to

social housing and other housing schemes as the wider UK population.

- Children and caring responsibilities: information about dependent children, family types, marital status and caring responsibilities of veterans to further understand the contribution of veteran households.
- Economic activity and employment: information on economic activity, socio-economic classification, working hours, occupation, place of work, industry and employment history.
- Geographical information: Home postcode, Lower Super Output Area (LSOA) and Middle Super Output Area (MSOA) in order to extract required locality information.
- Other demographic information: information about the veterans including explicitly:
 - i. Ethnicity
 - ii. Adult life stage
 - iii. Country of birth
 - iv. General health
 - v. Limiting long term illness (a proxy for disability)
 - vi. Household with limiting long term illness
 - vii. Highest level of qualification

Once ONS have completed the linkage, ONS would produce the file at LSOA, and additionally include the following geographies to the dataset as requested by MOD: Local Authority District, Unitary Authority, Clinical Commissioning Groups, Local Enterprise Partnership, region, country breakdown, and UK. All geographies will be provided subject to Statistical Disclosure Control rules and categories may be merged or otherwise remediated if necessary. This linkage will enable MoD and ONS to:

- establish whether VLD data quality supports future administrative data linkage projects; and
- develop an evidence base to identify demographic characteristics of veterans population at low level geographies in order to realise the benefits of this research in compliance with the Armed Forces Covenant.

ONS will serve as a trusted third party on behalf of the ADRN and will explore the feasibility and quality of linkage of the four datasets.

A linkage method has been designed based on creating bespoke matchkeys made up of various parts of the identifiers (surname, previous surname, forename (s) or initial, date of birth and gender). These have been constructing to allow for different errors to be allowed for while still being able to make a confident match (e.g. incorrect or missing forename on either Census or VLD).

ONS expects that administrative databases will include errors (e.g. initials instead of full names, Optical Character Recognition, errors, omissions and so on). The match rate will be dependent on the number of available correct matches, which is subject to records in the VLD not appearing in the 2011 Census (e.g. individuals not responding to the 2011 Census, not resident in the UK, death prior to 2011 Census day). Lessons learnt from the success of the matching methodologies and tools will be kept for future reference (and shared between the project collaborators to inform best practice in data linkage). Match rates and any information known about the quality of the links will be made available to researchers for understanding of the quality of the data linkage.

In future research it would be desirable to utilise death registrations data to identify deceased respondents, and data such as Patient Register data to update address information and improve accuracy. This research is outside the scope of this proposal.

The resulting dataset from this linkage will be de-identified and stored in the VML. Approved

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

researchers from the MoD will analyse this dataset to create a statistical report which will be released under the Statistics Code of Practice on Gov.uk (following approval from MoD Permanent Under Secretary and the Minister for Defence, Veterans, Reserves and Personnel (Min DVRP)).

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

The armed forces covenant (AFC) sets out the relationship between the nation, the government and the armed forces. It recognises that the whole nation has a moral obligation to members of the armed forces and their households, and it establishes how they should expect to be treated.

The covenant's 2 principles are that:

- The armed forces community should not face disadvantage compared to other citizens in the provision of public and commercial services
- Special consideration is appropriate in some cases, especially for those who have given most such as the injured and the bereaved

The covenant exists to redress the disadvantages that the armed forces community may face in comparison to other citizens, and to recognise sacrifices made. In developing our understanding of veteran's information needs it has also become clear that we need to identify household members as this group are also covered by the Community Covenants but are often overlooked which can result in social isolation, disrupted health and social care treatment, disrupted education for children, and difficulty for services in sharing data for safeguarding activities.

It is anticipated that the project will help improve existing services and help develop new services for veterans in England, Wales, Scotland and Northern Ireland. For example, a number of organisations are seeking demographic and location information on veterans for the purposes of ensuring appropriate commissioning and improved delivery of services. For example, in England the NHS, through Clinical Commissioning Groups (CCG), commission healthcare services for veterans and their households.

Overall, the successful delivery of this project will enable public and voluntary organisations to deliver improved services as described in the Armed Forces Covenant. The delivery of these services advocates a clear public benefit towards a specific population group (veterans and veteran households) and depends on a comprehensive evidence base which will be provided by this project.

Specifically, this project is expected to provide accurate information at a specific level of granularity of geographic information to improve existing services and develop new services for veterans in England, Wales, Scotland and Northern Ireland. An example is providing a wide evidence base to NHS via Clinical Commissioning Groups to develop targeted services to veterans and their households. GPs have expressed interest in understanding the number of veterans in their areas to target resources and care more appropriately. (e.g. referral to bespoke NHS services, charitable health and welfare providers with expertise in veterans).

This information will be invaluable in policy making by public health organisations that need to understand the costs, benefits and levels of service demand. This project will enable them to decide on the most effective means to plan, deliver, manage and promote services,

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

considering economic, public health and societal factors at a national and local level. Apart from direct benefits to the public it will indirectly result in the optimal use of resources of public services.

Providing information to organisations from the voluntary sector, such as charities, provides an additional public benefit towards supporting veterans and their households. It also expands their capabilities to deliver services more effectively especially at a local level.

Including veteran households is also an important aspect of this project as it offers more information on supporting veterans within societal norms as well catering for the needs of veteran household. Veteran households contribute significantly in supporting veterans and hence services should, according to the community covenant, be developed with them in mind as well. This enhances the anticipated benefits of this project by expanding the scope of the intended interventions.

The exploration of new methodologies and sharing best practice when linking datasets is anticipated to produce benefits in improving the methods currently employed and serve as a point of reference in future projects. This will greatly improve the capabilities of all contributing organisation in delivering their services to the public.

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

i. The use of data has clear benefits for users and serves the public good.

The research presents a clear public benefit in terms of providing information to government, local authorities, public health functions and voluntary organisations for the provision of locally targeted services and the improvement of existing services. It also examines different ways of linkage which might offer an improved method in future data linkage projects.

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 granularity of information at a low geographical level required for the linkage introduces a small risk of breaching the anonymity of the data subjects. Direct identifiers won't be made available to researchers but there is a risk of possible re-identification. Statistical disclosure control will be applied to eliminate the potential of disclosures before results are released from the VML. Researchers will be made aware of the risks of processing information about a person they might know. No consent is required by the data subjects for this project see section iv. Safeguards are put in place to mitigate risks of re-identification using the datasets in this research paper and in future linkage projects

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.

Safeguards are in place to make sure that any risks introduced by the proposed linkage are mitigated. These include a privacy impact assessment which has been undertaken by MoD, compliance with VML guidelines and statistical disclosure controls as set out in section B3. All processes will have adequate human control, in accordance with data security arrangements.

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 proposed data linkage is based on the legal framework set via the Statistics and Registration Services Act and the Data Protection Act. No additional legal gateways are required.

v. The views of the public are considered in light of the data used and the perceived benefits of the research.

There is no need for direct public engagement however there is increased interest from the Parliament, government departments and voluntary sector organizations to obtain more information on veterans and their households in support of the Armed Forces Covenant. The MOD and ONS have sought the views of a wide range of external stakeholders, including the charitable sector. In responding to a number of requests to the MOD from Freedom of Information requests, Parliamentary Questions and ministerial correspondence we have signaled our intent to provide more definitive information on veterans.

vi. The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public.

The research outputs will be clearly communicated to specific target audiences and the tools and methodology developed will be openly shared with the wider statistical service as set out in section B7.

B7 How will the findings of the research be disseminated?

The MoD and ONS researchers will review the success of the data linkage project. At the very least a two-page plain English report summary of the project, the dataset used, and the research question with an explanation of why the data was not robust enough to present the statistical findings will be published by the ONS and the MoD on Gov.uk. If the data linkage is successful and it is possible to present further meaningful statistical analysis the MoD will seek permission from the Permanent Under Secretary and the Minister for Defence for Veterans, Reserves and Personnel (Min DVRP) to release the findings as an experimental statistic in line with the statistics code of practice. This will ensure that all the statistical information is made available at the same time to all the interested parties including other government departments, the charitable sector and the public. If the linkage is successful the data will be kept for a certain period in which approved researchers will be able to apply to access the data in the VML to ensure that the public benefits of these data are fully utilised.

It is likely that the published information will be used to support further reports such as the

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

MoD's annual Armed Forces Covenant Report.

The outputs produced will also include research and methodology exploring the possibility of linking these datasets, and a discussion of the quality of the linking, to inform decision making around the use of administrative data to replace parts of the next Census in England, Wales, Scotland and Northern Ireland. ONS plan to publish a report assessing how much this dataset meets user needs as part of the evidence pack ahead of the Census White Paper and a recommendation on whether to include a census question on the Armed Force community.

With regards to the findings around linkage, these will be communicated by the relevant statistics services such as the Royal British Legion, the NHS, LAs and other government departments and to the wider community via their web pages and via seminars and workshops.

In summary, the results of the project including all research outputs will be made available in the public domain, and publication will be in line with ADRN policy.

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.

As part of accessing the linked data resource through the ADRN, this full linked de-identified dataset will not be available for use by other researchers at a micro level. However the dataset will be retained for 12 months to allow additional linkage to take place if necessary, or until new ADRN policy allows.

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*

veterans living in UK
veteran households

Justification for focusing on these subsections or groups:

Growing demand for information on veterans and their households to help improve existing services and help develop new services for veterans and their households in the UK.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

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

ONS, MoD, NRS and NISRA are data suppliers. ONS will provide the TTP data linkage service on behalf of the ADRN.

ONS will match, link and de-identify the data belonging to NISRA, NRS and MoD, as a service to those bodies, using its powers in section 22 of the Statistics and Registration Service Act 2007. The provision of data to provide this service does not constitute a lawful disclosure and so no legal gateways are required.

The Census elements of the de-identified research dataset will be made available to the MoD researcher(s) using the approved researcher gateway in s.39 SRSA. After consultation with the legal services no gateway is required to make the MoD elements available to the MoD researcher.

The MoD have produced a Privacy Impact Assessment (PIA) which has been signed off by the data security officer for MoD Head of Defence Statistics and Corporate Systems, the data protection officer for MoD Defence Business Services (the MoD Information and Records Management Lead) and the head of MoD-Veterans UK (the data owner).

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

Not applicable

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.
----	----	----	----	----	----	----	----	----	----	-----

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(17)09

LFS: Project to assess the representativeness of the Labour Force Survey Phase I

Purpose

1. The paper presents a proposal from the ONS Social Survey Division to use administrative data from Department of Work and Pensions and HM Revenue and Customs to assess the representativeness of the Labour Force Survey (LFS).

Recommendations

2. Members of NSDEC are invited to consider the 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 the proposal be resubmitted to a future meeting once implemented; or
 - iv. reject the proposal advising it be stopped from proceeding.

Background

3. The LFS is a study of the employment circumstances of the UK population. It is the largest household study in the UK and provides the official measures of employment and unemployment.
4. Government departments use the results of the study to identify how and where they should be using public resources. They use the information to check how different groups in the community are affected by existing policies and to inform future policy changes. The European Union uses the results to decide how much funding they will give the UK to improve local area employment prospects, facilities and opportunities.
5. Falling response rates have led to an increased focus on understanding if the LFS sample is still representative. In 2012 ONS used 2011 Census data with the corresponding period from the LFS data, to assess the LFS and issue any adjustments to account for bias from non-responding households. No adjustments were identified at that time. Since then the survey has experienced falling response rates.
6. The ONS Social Survey Division are looking to do the analysis outlined in this application to assess the current representativeness of the LFS. This analysis will be performed using data from up-to-date administrative sources from the Department for Works and Pensions and HM Revenue and Customs. These data sources are expected to have reasonable and up-to-date coverage of the whole labour market.
7. The main outcomes of this phase of the project will be used to improve the representativeness of the LFS. Improving the robustness of the LFS will improve the accuracy of the LFS data that is available to government departments and relevant stakeholders to enable more informed policy making.

Petros Saravakos, NSDEC Secretariat, Central Policy Secretariat, UK Statistics Authority, 10 April 2017

List of Annexes

Annex A: Application: LFS: Project to assess the representativeness of the Labour Force Survey Phase I, Christopher Daffin

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.
----	----	----	----	----	----	----	----	----	----	-----

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

--

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

Section A
Application Details

A1	Responsible Owner									
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; padding: 5px;">Full Name: [REDACTED]</td> <td style="width: 50%; padding: 5px;">Position: [REDACTED]</td> </tr> <tr> <td style="padding: 5px;">Address: [REDACTED]</td> <td style="padding: 5px;">Email: [REDACTED]</td> </tr> <tr> <td></td> <td style="padding: 5px;">Telephone: [REDACTED]</td> </tr> <tr> <td></td> <td style="padding: 5px;">Organisation: [REDACTED]</td> </tr> </table>			Full Name: [REDACTED]	Position: [REDACTED]	Address: [REDACTED]	Email: [REDACTED]		Telephone: [REDACTED]		Organisation: [REDACTED]
Full Name: [REDACTED]	Position: [REDACTED]									
Address: [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>										
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; padding: 5px;">Print Name: [REDACTED]</td> <td style="width: 50%;"></td> </tr> <tr> <td style="padding: 5px;">Signature:</td> <td></td> </tr> <tr> <td style="padding: 5px;">Date: [REDACTED]</td> <td></td> </tr> </table>			Print Name: [REDACTED]		Signature:		Date: [REDACTED]			
Print Name: [REDACTED]										
Signature:										
Date: [REDACTED]										

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

A2 Applicant Details (if applicant is not the responsible owner)

Full Name	Position
Address:	Email:
	Telephone:
	Organisation:

A3 Project Information

Project Title: Project to assess the representativeness of the Labour Force Survey	
Start Date: 10 March 2017	End Date: 31 July 2017
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)
Department for Work and Pensions HM Revenue and Customs Office of National Statistics (Population Methodology and Statistical Infrastructure Division, Data as a Service Division, Labour Market and Households Division, Survey Methodology and Statistical Computing)	Legal gateways have been established at the time of the application to obtain the data ONS Legal Services have advised formalising this agreement by contacting data suppliers to confirm this work is in line with the agreement for provision of the data

Division, Health Analysis and Life Events
Division, ONS Social Surveys Division)

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 (please specify).....

Is this a secure site?

- ☒ Yes ☐ No

Section B
Project Details

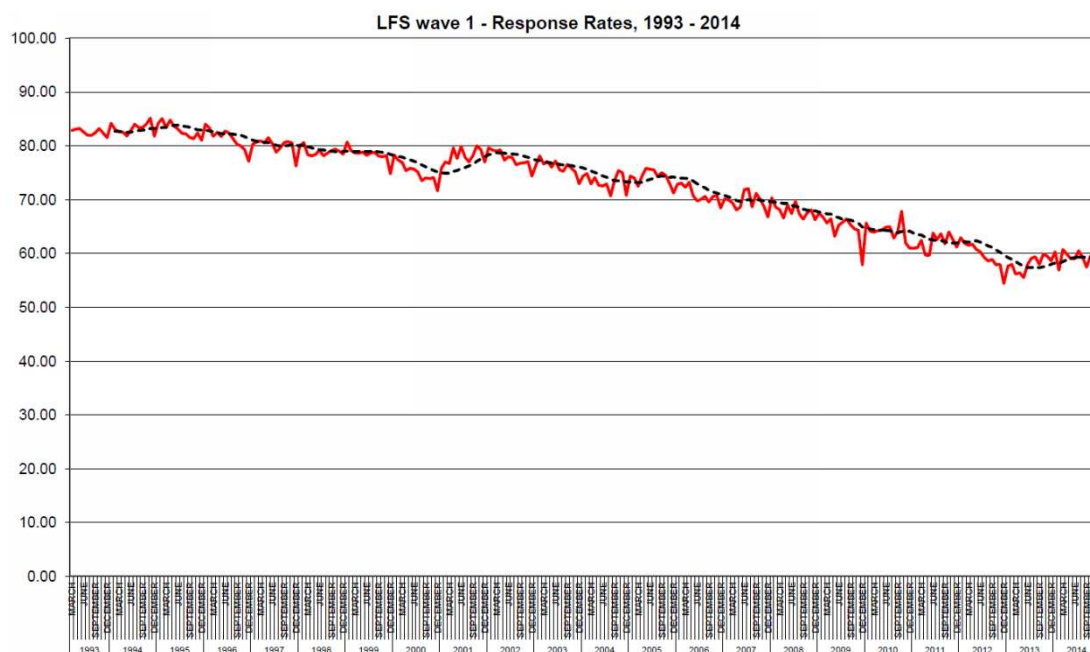
B1 **Please provide a brief high level summary of the research giving necessary background** (max 500 words)

The Labour Force Survey is a study of the employment circumstances of the UK population. It is the largest household study in the UK and provides the official measures of employment and unemployment.

Government departments use the results of the study to identify how and where they should be using public resources. They use the information to check how different groups in the community are affected by existing policies and to inform future policy changes. The European Union uses the results to decide how much funding they will give the UK to improve local area employment prospects, facilities and opportunities.

In 2012 ONS linked 2011 Census data with the corresponding period of the LFS to assess if the LFS was representative and if an adjustment was required to account for any bias from non-responding households. The study concluded that no adjustment was necessary. Since then response on the LFS has continued to decline to such an extent that there is concern that the survey results could now be biased, leading to incorrect conclusions from the survey results.

Certain factors, in particular falling response rates, have led to an increased focus on understanding if the Labour Force Survey (LFS) sample is still representative. The National Statistics Executive Group has requested that work be undertaken to understand this.



In order to assess the current representativeness of the LFS ONS Social Survey Division would like to do similar analysis to the 2011 Census linkage work but using more up-to-date administrative data or possibly wider data sources (e.g. Big Data). This project will use data sets from the Department for Work and Pensions and HM Revenue and Customs to investigate their feasibility in assessing the representativeness of the LFS. These data sources are expected to have reasonable and up-to-date coverage of the whole labour market and should therefore provide useful information for assessing the representative of the LFS. Particularly on those households that would not or could not respond to the LFS survey.

The outcomes of this project will be used to inform about the representativeness of the Labour Force Survey and not to make adjustments to account for any issues found. This presents a clear public benefit in providing accurate and timely information to government departments and relevant stakeholders to enable informed policy making.

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>		DWP/HMRC Benefit and Income Data Source 2011-16 DWP/HMRC Customer Information System 2011-16		
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>		ONS Labour Force Survey ONS Annual Population Survey		
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?
(max 500 words)

Both the LFS data and the required variables from the Customer Information System (CIS) and Benefits and Income Data (BIDs) will be processed and stored in the ONS Statistical Research Environment. This is a CESG¹ accredited environment, which requires hashing (a one-way encryption process that is irreversible) of identifiable data.

¹ UK Government's National Technical Authority for Information Assurance (CESG)

To minimise the risk of identifying individuals and their attributes while retaining human oversight of the process, separation of duties is implemented. Specifically, staff involved in the de-identification (and encryption) of the data, and have access to identifiable information will not have access to attribute or linked data.

In terms of physical security, the environment where data will be processed and stored is physically disconnected and air-gapped from other ONS infrastructure. Access is allowed only to authorised staff and no data will be removed at any stage from the site.

All outputs (final and intermediate) produced within the environment are statistically disclosure controlled before being exported. This ensures that only anonymised aggregate data may leave the secure environment upon appropriate authorisation.

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

LFS data from responding households and people will be linked to the DWP Benefit and Income Data Source (BIDS) and Customer Information Service (CIS) data sets. In order to provide information on the LFS non-responding households the address of non-responding households will be linked to the address on the BIDS and CIS data sets. Aggregate measures for responding and non-responding households will be compared to assess if those measures differ between the two groups.

ONS Data as a Service Division (DaaS) will be responsible for the data linking and providing data for Survey Methodology and Statistical Computing Division to do the analysis. ONS Social Survey Division and Labour Market and Households Division will quality assure the resulting analysis.

All data sets will be geocoded by DaaS to add Unique Property Reference Number (UPRN) based on the address details in the data. LFS will be linked to CIS by UPRN and then CIS to BIDs by encrypted National Insurance Number.

There will be two parts to the comparison of the LFS and BIDS data sets. The first will produce distributions of key BIDS variables from the full BIDS data set for each quarter up to the latest time period available. This does not involve the use of linked data and will be for aggregated data only. We will compare these distributions with the weighted LFS corresponding distributions. Variables of interest for this section are:

- CXBENTYP – benefit type
- ISLP – Income Support lone parent flag
- CCTEN – tenancy type

The second part will use the linked data set to compare aggregate measures from the LFS for responding households with aggregate measures calculated from the same households in the BIDS data set. Aggregate measures from the LFS will also be compared to aggregate measures calculated from the BIDS data set for those addresses that did not respond to the LFS. We will compare the distributions of responders and non-responders with respect to a selection of BIDS variables. First the CIS data set will be linked to the LFS survey period at address level, including from the CIS data set every person at each linked address. Next the data set will be linked to the BIDS data that relates to the LFS time period being studied by quarter. A comparison will be made between aggregate LFS and BIDS information for unemployment benefit, income support, pension and a proxy for employment. The linked data set will not be used to compare an individual's response to the LFS

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

survey.

The following flags will be derived from the BIDs data:

- Unemployment benefit – where an individual has a benefit type of Job Seekers Allowance (JSA)
- Income support – where an individual has a benefit type of Employment and Support Allowance (ESA)
- Pension – where an individual has a benefit type of Retirement Pension (RP)
- Employee proxy – where an individual has an entry in the Pay As You Earn (PAYE) data set

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

Concern has been raised from users of the LFS about the extent to which the LFS published results are representative of the population the survey is sampled from. This could have implications for the accuracy of the outputs from the survey, including key measures such as the employment and unemployment rates which could have significant impacts on Government policy decisions.

The linked data set could also provide information on the feasibility of using administrative data, for example to replace some of the questions asked to respondents in the LFS. This would fit with work planned as part of the Data Collection Transformation Programme.

Public benefit for this piece of work is to help optimise surveys, reducing completion times and increase response rates. The Labour Force Survey (LFS) is a study of the employment circumstances of the UK population. It is the largest household study in the UK and provides the official measures of employment and unemployment.

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 use of data has clear benefits for users and serves the public good.***
Improvement of the quality of the Labour Force Survey has clear benefits for the public good in terms of better, more reliable statistics leading to better public policy decisions by government.
- The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is appropriately considered.***
Data will be pseudonymised prior to linkage for the specific purpose of protecting data subjects' identity. Safeguards, such as segregation of duty, are in place to ensure that the anonymity of data subjects is ensured.
Only statistical summaries of the data will be used in comparisons between the LFS and the administrative data sources. Addresses for the LFS are sampled from a list of addresses compiled by the Post Office. A letter and information leaflet are sent to all of these sampled addresses informing the occupant about the survey, where to find additional information and that the information they supply will be used to produce

	<p>statistics that will not identify anyone in the household. The leaflet informs the occupant that the survey information is also provided to other approved organisations for statistical purposes only. The only information regarding non-responders available in this project is their address, which was part of the LFS survey design. Only statistical summaries for non-responding addresses will be used in comparisons between the LFS and the administrative data sources.</p>
iii.	<p><i>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.</i></p> <p>All staff working with these data are (at least) Security Cleared and trained. Processes put in place ensure separation of duty preventing anyone seeing identifiable data once the data are linked and human oversight is ensured throughout the project activities.</p>
iv.	<p><i>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.</i></p> <p>ONS Legal services have confirmed this is legal. The data owners have been advised of the project and are content. This is reflected in the MoU between the collaborating departments. The proposed data linkage is compliant with the Statistics and Registration Service Act 2007 as well as the Data Protection Act.</p>
v.	<p><i>The views of the public are considered in light of the data used and the perceived benefits of the research.</i></p> <p>The views of the public have not been sought with regards to the use of the data in this piece of research, nonetheless there is a clear public benefit for the linkage (as described in section B5).</p>
vi.	<p><i>The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public.</i></p> <p>The project outcomes will be published on the ONS website and will be made available to the public. Any new methodology or tools that might be developed will be shared within ONS.</p>

B7	How will the findings of the research be disseminated?
	<p>A report will be published on the ONS web site. The report will include summaries of the factors relevant to representativeness of the LFS. No information will be included that could be used to identify individuals or businesses following established statistical disclosure controls.</p>

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.
	<p>If the project finds that linkage of LFS with BIDS data is useful then the exercise could be repeated in order to monitor the ongoing representativeness of the LFS.</p> <p>Many ONS surveys are seeing declining response rates. The outcomes from this project could be</p>

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

used to inform any measures taken to mitigate this.

The linked data set may also be useful to inform the Data Collection Transformation Programme. In particular around the feasibility of replacing income questions on the LFS and other surveys.

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:

Justification for focusing on these subsections or groups:

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

The data owners DWP/HMRC have been advised of the project and are content. ONS Legal Services have confirmed the legality of the use of the data for this purpose, subject to DWP/HMRC agreement. This is reflected in the MoU between the collaborating departments.

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

HMRC and DWP data will be obtained using the established legal gateways (MoU) between the different collaborating departments.

UK Statistics Authority
National Statistician's Data Ethics Advisory Committee

NSDEC(17)10

Considered via correspondence

Microdata Release Panel: The societal impacts of water problems (such as flood and supply interruption) on health and wellbeing.

Purpose

1. This paper presents a proposal by the Microdata Release Panel for use of ONS data by Simetrica Ltd, a research consultancy specialising in policy evaluation, who have been commissioned by Thames Water Ltd, to carry out research into the societal impacts of water problems on health and wellbeing.
2. This proposal was approved by NSDEC via correspondence and is presented here to transparently record this correspondence.

Background

3. The Microdata Release Panel (MRP) governs access to the secure ONS Virtual Microdata Laboratory, and provides approved researchers access to de-identified ONS data in order to undertake approved projects.
4. In 2016, following a public consultation of the Approved Researcher process, it was agreed that where necessary, the MRP would refer applications to use ONS data via the approved researcher gateway to NSDEC for ethical review.
5. In the NSDEC meeting on 24 January 2017 the Committee were informed that some of these proposals would need to be considered via correspondence in order to ensure equality in timeliness of access to ONS data.
6. In the discussion that followed, members agreed that:
 - i. it would be most appropriate for all members to be sent the application for review;
 - ii. a sufficient number of members would need to respond in accordance with a quorate meeting for any proposal being considered by correspondence to be formally approved; and
 - iii. where expedited review is sought and NSDEC approve the proposal, the application and summary of member's comments will be published as part of NSDEC's next meeting.
7. The secretariat is working with the MRP secretariat to minimise the number of projects that need to be considered by correspondence in the future.
8. The project presents a clear public benefit to improve the understanding of the impacts of water-related risks to the health and wellbeing of the public as well as provide more robust methods to infer the wants and needs of customers. The research will also support Thames Water in achieving their vision of zero harm and improve health and wellbeing, for their employees, customers and everyone that comes into contact with the company. It will also help them to better understand and respond to the needs of its customers. It offers an opportunity to provide an evidence base for service planning and

decision making for water companies, regulators and relevant public organisations, in relation to the health and wellbeing and water management in the UK.

Petros Saravakos, Central Policy Secretariat, UK Statistics Authority, 28 February 2017

List of Annexes

Annex A: Application: The societal impacts of water problems (such as flood and supply interruption) on health and wellbeing

Annex B: Summary of discussions and actions via correspondence

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

Annex A: Minutes of correspondence with regards to the application, 11 April 2017

Minutes of Correspondence 11/04/2017

Participating Members

Mr Ian Cope (Chair)
Mr Robert Bumpstead (Deputy Chair)
Ms Vanessa Cuthill
Mr Keith Dugmore
Ms Annie Hitchman
Mr Colin Godbold
Ms Isabel Nisbet
Ms Marion Oswald

UK Statistics Authority

Mr Petros Saravakos

Office for National Statistics

Mr Pete Stokes (Microdata Release Panel)

Apologies:

Dr Dean Machin
Mr Osama Rahman
Mr Neil McIvor
Professor Martin Severs

1. Public good and user benefit

- 1.1. Members agreed that the proposed research has a clear public benefit.
- 1.2. Queries were raised to whether the application would benefit by linking the Thames Water health and wellbeing strategy to any imposed / agreed government objectives or schemes. The initiative could also feed into the vulnerable customers scheme to enhance the public benefit of the application and strengthen the connection between health, prices and public benefits.
- 1.3. Members heard that the research would contribute to the vulnerable customers scheme by considering the impact of sewage odour on up to three vulnerable groups (such as those on low incomes, those aged over 65, and those who are unemployed). Ofwat may decide to consider this research for the next price review by including a special price reduction for the affected customers. Researchers will identify the exact vulnerable groups to consider based on research conducted by Ofwat and recommendations by Thames Water.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

2. Data, methods, confidentiality and security

- 2.1. Members reached agreement on more clarity required on the exact nature of the water-related problems referred. Mr Stokes clarified on behalf of Simetrica that the project will research the impact of smelling sewage odour on health and wellbeing. The application will be amended to reflect this particular research scope.
- 2.2. Queries were raised with regards to the link between the risk and the health and wellbeing of Thames Water customers. It is not clear where health data are collected from and how the impact of incidents on health and wellbeing will be measured. Mr Stokes informed members that subjective health status variables from the ONS Annual Population Survey will be used to examine the impact of sewage odour linking the two aspects.
- 2.3. Members identified that there is a small risk of harm (financial) to customers as a result of the research. Researchers informed members that the prices for water provision and waste treatment are regulated and not decided by Thames Water. As a result there is possibility that the research outcomes may be used to influence the decision of Ofwat during the next price review.
- 2.4. Further information on the timeframe of the data used to delineate the study group which is not specified in the application is requested. This might impact on the quality / validity of research. It is advised that more information is required:
 - i. to ensure the timelines of health and wellbeing information; considering that incidents should occur before the surveys are conducted.
 - ii. to distinguish populations affected by an incident but were not living in the area when the survey was undertaken
- 1.1. Researchers indicated that geographical information on sewage treatment works and the time of their operations will be used along with geographical information at the time of the Annual Population Survey.

3. Transparency in the use of data

- 3.1. Members agreed that the application could benefit by further explaining the relationship between Simetrica and Thames Water. It is a common practice for commercial organisations to employ research consultancies, but it provides little transparency of who holds the ownership of the data and research outcomes.
- 3.2. Mr Stokes clarified on behalf of the responders that only the Simetrica project team will have access to the ONS in the VML environment. Simetrica and Thames Water will share ownership of the cleared aggregate research outcomes. Thames water will retain ownership of the sewage treatment works data which is used in the analysis, but may share it with Simetrica for any future applications.
- 3.3. Simetrica provided assurances that research outcomes will be published in the public domain as open access academic journals, on the Simetrica and also on the ONS website.

- 3.4. Members were informed that due to legal reasons Thames Water will retain the right not to put their name on the on academic publications or publications on Simetrica's website. Nevertheless, the robust outputs of this research will be reflected in Thames Water health and wellbeing strategy which will be available to regulatory and supervisory authorities such as Ofwat. .

4. Conclusions

- 4.1. The Microdata Release Panel approved this project and referred to NSDEC for ethical consideration via correspondence
- 4.2. All participating members agreed that the project demonstrated clear public benefit.
- 4.3. Members agreed that the project could proceed subject to the scope of the project, as detailed in 2.1 being made clearer and reflected in the application.

**Petros Saravakos, NSDEC Secretariat, Central Policy Secretariat, UK Statistics Authority,
10 April 2017**

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.
----	----	----	----	----	----	----	----	----	----	-----

Annex B: Application: The societal impacts of water problems (such as flood and supply interruption) on health and wellbeing, 7 March 2017

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

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

Section A
Application Details

A1 Responsible Owner

Full Name: [REDACTED]	Position: [REDACTED]
Address [REDACTED]	Email: [REDACTED]
	Telephone: [REDACTED]
	Organisation: [REDACTED]

Declaration to be signed by the responsible owner

I have met with and advised the applicant on the ethical aspects of this project design (*applicable only if the responsible owner is not the Applicant*).

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.

I am satisfied that the research complies with current professional, departmental and other relevant guidelines.

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.

I will provide notification when the study is complete if it or fails to start or is abandoned.

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.

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.

Print Name: [REDACTED]
Signature:
Date [REDACTED]

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

A2 | Applicant Details (if applicant is not the responsible owner)

Full Name		Position:	
Address:		Email:	
		Telephone:	
		Organisation:	

A3 | Project Information

Project Title: The societal impacts of water problems (such as flood and supply interruption) on health and wellbeing.

Start Date: 10 March 2017

End Date: 31 August 2018

Project Sponsor (select all that apply)

☐ ONS

☐ ADRN

☐ GSS

☐ Collaboration

☒ Other (Please specify).....Thames Water <https://www.thameswater.co.uk/>

A4 | Collaboration and Sponsors

List of Collaborators/Sponsors	Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)
Thames Water Utilities Ltd	

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 (please specify).....

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)

Thames Water Utilities Ltd, known as Thames Water, has commissioned Simetrica Ltd, a research consultancy specialising in policy evaluation, to carry out research into the societal impacts of water problems on health and wellbeing.

Thames Water is the private utility company responsible for the public water supply and waste treatment in large parts of Greater London, the Thames Valley, Surrey, Kent, Wiltshire, Gloucestershire and some other areas of the United Kingdom. It is the UK's largest water and wastewater services company and its 15 million customers comprise 27% of the UK population.

The research outcomes will contribute to Thames Water's health and wellbeing strategy which sets out the steps they need to take to achieve their vision of zero harm and improve health and wellbeing, for their employees, customers and everyone that comes into contact with the company. It also informs the price review 2019 process by providing triangulation to traditional methods used by water companies to infer the wants and needs of customers.

[Simetrica](#) is a private sector consultancy with expertise in the ethics of policy evaluation, non-market evaluation, causal inference, cost-benefit analysis and behavioural science. They work with a wide range of high-profile organisations in the public, private and not-for-profit sectors, including the United Nations, OECD, UK government departments, Government of Australia, Government of Ireland and Government of Canada. UK government clients have included the Airports Commission (Airport expansion consultation) and the Department of Culture, Media and Sport on the wellbeing benefits of sports and culture. These analyses were peer reviewed by government and independent academics, and have been subsequently published in academic journals.

The purpose of the research project is to:

- Quantify and model the impact on individual and societal health and wellbeing from water-related risks (flood and supply interruption) to help develop strategies to address these;
- Test the hypothesis of whether reductions in water-related risks and outcomes are associated with improvements in health and wellbeing for those living in affected areas;
- Identify the negative consequences for customers impacted by water-related incidents and analyse relevant demographic trends to help understand whether particular societal groups (such as those with low income and/or low educational attainment) are at disproportionate risk;
- Help identify and define trends which relevant water management and government organisations need to consider and mitigate in their service planning and decision making to assist those members of public at risk; and
- contribute to Thames Water's health and wellbeing strategy which sets out the steps they need to take to achieve their vision of zero harm and improve health and wellbeing, for their employees, customers and everyone that comes into contact with the company.

The project will analyse the wealth of data collected by water companies and regulatory bodies in the UK, in order to model the health and wellbeing impacts of water-related risks. Initially, de-identified health and wellbeing data from the ONS Annual Population Survey will be linked to data held by Thames Water on water disruption events (e.g. historic and real-time data on water catchment management, such as flooding, odour and supply interruptions) at postcode level, and with controls for known socio-demographic and socioeconomic drivers of health and wellbeing. The data at postcode level will be de-identified and cleaned to prevent any possible disclosure of data subjects or individuals. Depending on how this initial research goes, it may translate into further research sponsored by Thames Water data and potentially other water companies operating in other geographic areas in the UK. This will involve a similar data matching procedure.

This study will develop on a number of previous academic journal papers which have assessed the impact of water-related incidents on wellbeing (e.g. Frijters et al.'s (2009) assessment of the impacts of droughts on wellbeing in Australia using the Household Income and Labour Dynamics Australia survey (HILDA) data set (published in the [Journal of Population Economics](#) 22(2):445-461)). This will be the first research to explore the impact of

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

water-related risks on societal health and wellbeing using best practice quasi-experimental methods in the UK. Previous studies have not used incident data from water companies and regulatory bodies, so this study will represent the first of its kind in the UK.

An output from the research will be a high-level summary report which will quantify the impact on individual and societal health and wellbeing from water-related risks (such as flood and supply interruption). All outputs and tables in the report will be checked by the ONS Virtual Microdata Laboratory (VML) statistical support team to ensure that there is no possible disclosure of data subjects. The research report will enable water companies and the economic regulator for the water sector in England and Wales, Ofwat, to understand better and address the impacts of their sector on the health and quality of life of UK citizens and guide future policy decisions and service planning in relation to water management and health and wellbeing. There will also be high-level academic journals produced with non-disclosing data table and maps.

The ONS Microdata Release Panel (MRP) approved the proposal at its meeting on 20 January 2017 on the grounds that there was a legal gateway to access the data, it was appropriate use of ONS data and a public benefit was demonstrated (provides an evidence basis for public health and water management decision making and service planning).

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>				
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>			ONS Annual Population Survey data, 2011 – 2015. (Via the VML) Understanding Society - Special Licence data (2009-2015) (Application Granted through UK Data Service)	

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>			Water incident data from Thames Water	

B3 **How will information be kept confidential and data kept secure?**
(max 500 words)

Access to data will only take place within the secure ONS Virtual Microdata Laboratory (VML) environment and all outputs will be checked by the VML team prior to release to ensure disclosure control and the confidentiality of data subjects is protected. All analysis and use of the data will be within the VML at a secure VML setting at one of the ONS offices. The data will be linked within the VML environment. The Simetrica researchers have recently been accredited as ONS Approved Researchers following their successful completion of training on the Safe use of data in research environments (SURE) course. No data or report drafts will be shared with anyone outside of the project team and ONS will ask for sight of the report before it is published.

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

Water companies and the regulator record data on water incident events throughout the year. This includes detailed temporal data about water disruption, odour problems from sewerage works and flood events, logged at postcode level (and then cleaned to ensure that it does not contravene rule 4.1.2 of Microdata Handling and Security - Guide to Good Practice on data matching). This de-identified data will be imported into the VML where it will be linked with the health, wellbeing and other societal outcomes from the detailed Annual Population Survey wellbeing data from April 2011 – March 2015, at postcode/output area geography and then analysed to ascertain a quasi-experimental estimate of the impact on health and wellbeing in the postcodes considered.

Simetrica will use the Annual Population Survey data to assess the health and wellbeing of individuals in areas where there has been a historical incident of a water-related problem. They will compare the health and wellbeing of these individuals with a control group of individuals who were not living in the affected area at the time. They will control for initial differences across the affected and non-affected groups using a set of standard demographic variables available within the Annual Population Survey dataset (e.g. income,

labour force status, and education).

The Annual Population Survey **variables** requested are:

- Subjective wellbeing
 - Life satisfaction
 - Happiness
 - Worthwhile
 - Anxiety
- General health
- Social relations
- Labour Force variables
 - Income
 - Employment status
 - Marital status
 - Housing status
 - Dependent children
 - Education level
 - Age
 - Date of interview

Simetrica have also been granted access to use Special Licence [Understanding Society](#)¹ data which provides postcode output area level information through the UK Data Service within this research project. Simetrica plan to use fixed effects regression analysis and difference-in-difference analysis in the Understanding Society data which allows for longitudinal analysis as it tracks the same households over time. This will enable a comparison of results derived from models from different data sets (i.e. Annual Population Survey and Understanding Society) and methods, which will provide a more nuanced view on the magnitude and direction of association between water events and wellbeing. There will be no linking of the Annual Population Survey and Understanding Society data together.

The final and any intermediate outputs from the VML will be cleared by ONS to ensure the confidentiality of data subjects is protected. The initial output from the research is scheduled to be a high-level summary report elucidating the impact of certain water incidents on the health and wellbeing of individuals and society for Thames Water. Given the level of interest in the water industry, this is likely to be developed into a wide scale academic grey literature style publication which will be published on the Simetrica Website, and a link to the report will be included on the ONS website. It is also likely to be published on the website of the water companies that sponsor the full research as well as the Water regulator Ofwat. Any reports or academic papers produced will include data tables, thematic maps and charts to illustrate trends, but will not include any detail of individual incidents or any other information from which persons could be identified.

¹ <https://discover.ukdataservice.ac.uk/catalogue/?sn=7248&type=Data%20catalogue>

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

B5	Please outline the proposed benefits of the project (<i>max 500 words</i>)
-----------	---

The purpose of the research project is to:

- Quantify and model the impact on individual and societal health and wellbeing from water-related risks (flood and supply interruption);
- Test the hypothesis of whether reductions in water-related risks and outcomes are associated with improvements in health and wellbeing for those living in affected areas; and
- Help identify and define trends which relevant water management and government organisations need to consider in their planning and policies.

The research outcomes will help:

- Develop a better understanding of the types of water incidents and their impact on the health and wellbeing of society;
- identify and analyse relevant demographic trends to help understand whether particular societal groups are at disproportionate risk;
- Inform investment decisions and regulatory guidance surrounding the delivery of the water system to the benefit of UK consumers. For instance, the research can direct attention to how water companies can reduce the risk of flooding to match the impact that these extreme events have on UK resident's quality of life; and
- Inform thinking about solutions to protect UK consumers, e.g. to triangulate evidence collected by water companies using other methods, to ensure that they prioritise what UK residents most want and need (as determined by the impact on wellbeing).

These research outcomes should achieve a public benefit by providing an evidence base for service planning and decision making for water companies, regulators and relevant public organisations, in relation to health and wellbeing and UK water management. The research outcomes will contribute to [Thames Water's health and wellbeing strategy](https://sustainability.thameswater.co.uk/safe-sustainable-workforce/health-and-wellbeing)² which sets out the steps they need to take to achieve their vision of zero harm and improve health and wellbeing, for their employees, customers and everyone that comes into contact with the company, including local communities. It will also help them to better understand and respond to the needs of its customers. Simetrica plan to develop the outcomes from this research to inform further analysis on the impact of water-related incidents on health and wellbeing in other UK regions.

² <https://sustainability.thameswater.co.uk/safe-sustainable-workforce/health-and-wellbeing>

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*)

i. The use of data has clear benefits for users and serves the public good

This will be the first research to explore the impact of water-related risks on societal health and wellbeing using best practice quasi-experimental methods in the UK. Therefore, it will inform knowledge and understanding and/or challenge existing views about the impact of these events on society and the economy.

The research outcomes will help:

- Develop a better understanding of the types of water incidents and quantify and model the impact on individual and societal health and wellbeing;
- identify and analyse relevant demographic trends to help understand whether particular societal groups (e.g. people on low income or with low educational attainment) are at disproportionate risk;
- Inform investment decisions and regulatory guidance surrounding the delivery of the water system to the benefit of UK consumers; and
- Inform thinking about solutions to protect UK consumers and improve their health and wellbeing.

These outcomes will therefore achieve a public benefit by providing an evidence base for decision making and service delivery in relation to water management and public health and wellbeing. It will contribute to Thames Water's health and wellbeing strategy which includes local communities and customers. This will help water management, government, health and voluntary organisations plan more effective services and interventions, and hold the water companies to account over their performance against the strategy.

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 data (including the data from Thames Water and Understanding Society) will be accessed in the secure ONS Virtual Microdata Laboratory (VML) and is de-identified to protect the confidentiality of data subjects. The Approved Researcher scheme will be the legal gateway used to access the data and the researchers have successfully completed the Safe User of Research data Environments (SURE) training and have signed and will adhere to the Approved Researcher declaration setting out how they will manage the data and protect the confidentiality of data subjects in line with the Statistics and Registration Service Act 2007. The researchers will not be able to take the data outside of the VML and their outputs – both intermediate and final – will be checked and cleared by the ONS VML branch to ensure the confidentiality of data subjects is protected. ONS will request a copy of the final report prior to publication.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

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.

No new technologies are being used. The research methods employed will be openly available for further scrutiny or replication of results.

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.

Access to the potentially disclosive data will be in a secure environment (VML) and via an approved legal gateway (Approved Researcher scheme). This is in compliance with the Statistics and Registration Service Act 2007. The methods used are compliant with the principles in the Data Protection Act.

v. The views of the public are considered in light of the data used and the perceived benefits of the research.

Whilst the views of the public have not been sought with regards to the research, there is a clear public benefit for the analysis (as described in B6.1 above). A public consultation on the Approved Researcher scheme in 2015/16 recommended that commercial organisations should be allowed to access ONS research data where there is a clear public benefit.

vi. The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public.

The research methodology and outcomes will be made public and reported on. Simetrica aim to publish the research in peer-reviewed reports to inform and expand knowledge regarding the impact that the water sector has on the quality of life of UK residents. Simetrica will also submit papers to the top ranked academic journals, using Open Source publication guidelines. Simetrica plan to publish the research outcomes on their website as well as in literature produced by water companies and Ofwat. Initial results could be published in 2017 with a full academic submission/publication expected by mid-2018.

To help promote greater transparency and in compliance with the updated Approved Researcher scheme, the researchers have agreed to their details being included on a public record of Approved Researchers and to publishing the findings of their research (including on the ONS Approved Researcher pages). ONS will work with Simetrica and Thames Water to consider the inclusion of this analysis as a published case study on the ONS website setting out the research methodology and outcomes, and their impact on public service delivery and decision/policy making.

B7 How will the findings of the research be disseminated?

The results will be published by Simetrica – including publication in academic journals and on their website. They plan for the results to be made available on the websites of water companies and potentially the government regulator, Ofwat. A link to these will also be included on the ONS Approved researcher web pages in accordance with the Approved Researcher transparency criteria.

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.

None

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:

The project focuses on those individuals that are included in the ONS Annual Population Survey and live in areas that have been affected by water-related incidents, such as flooding. The research will also consider a controlled group of survey respondents who live in geographical areas not impacted by water incidents, to help compare their health and wellbeing. All ages, gender, ethnic groups that are included in the Annual Population Survey will be included in the analysis.

Justification for focusing on these subsections or groups:

To meet the scope of the project by focusing on the health and wellbeing of people and communities affected by water-related incidents

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

The ONS Approved Researcher scheme is the legal gateway being used to access the data. This is in compliance with the Statistics and Registration Services Act 2007.

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

Data from Thames Water will be used with their consent as the project sponsor. The special licence Understanding Society data use has been approved by the UK Data Service at the University of Essex which is the data owner.

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.
----	----	----	----	----	----	----	----	----	----	-----

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(17)11

Microdata Release Panel: Study of Small and Medium-sized Enterprises (SMEs) in the UK as part of the European Commission's SME Performance Review 2016/17

Purpose

1. This paper presents a proposal referred by the Microdata Release Panel for access and use of ONS data by London Economics, a policy and economic consultancy, to understand the Small and Medium Enterprises (SMEs) employment sector in the UK.

Recommendations

2. Members of NSDEC are invited to consider the 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 the proposal be resubmitted to a future meeting once implemented; or
 - iv. reject the proposal advising it be stopped from proceeding.

Background

3. The European Commission has commissioned London Economics to conduct the Annual SME Performance Review in the UK.
4. Small and medium-sized enterprises (SMEs) represent 99% of all businesses in the EU. The definition of an SME is important for access to finance and EU support programmes targeted specifically at these enterprises. The main factors determining whether an enterprise is an SME are staff headcount and either turnover or balance sheet total. The following are included. The following companies are included in the definition of small and medium sized enterprise.

Company category	Staff headcount	Turnover or	Balance sheet total
Medium-sized	< 250	≤ € 50 m	≤ € 43 m
Small	< 50	≤ € 10 m	≤ € 10 m
Micro	< 10	≤ € 2 m	≤ € 2 m

5. The SME Performance review is one of the main tools the European Commission uses to monitor and assess progress in implementing the Small Business Act on an annual basis in each country. The Small Business Act (SBA) is an overarching framework for the European Union policy on SMEs. It aims to improve the approach to entrepreneurship in Europe, simplify the regulatory and policy environment for SMEs, and remove the remaining barriers to their development.
6. Besides reviewing recent trends and the short-term outlook for SMEs, the 2016/17 report will also contain a chapter focusing on self-employment. This will focus on employment creation by enterprises created originally as self-employed or one-person enterprises. It will also include aggregate figures for net employment creation by firm type, sector and firm size.

7. The proposal will see London Economics analyse ONS de-identified Business Structure Database data within the ONS secure Virtual Microdata Laboratory. The Executive Agency for Small and Medium-sized Enterprises (EASME) has endorsed this project and has urged national statistical organisations to offer a positive response to London Economics.
8. This project aims to improve the understanding of the SME sector in the UK and provide in depth information on SME employment creation dynamics, profitability and export performance. This creates a comprehensive evidence base for national policy making and informing key government initiatives.

Petros Saravakos, NSDEC Secretariat, Central Policy Secretariat, UK Statistics Authority, 10 April 2017

List of Annexes

Annex A: Application: Study of Small and Medium-sized Enterprises (SMEs) in the UK as part of the European Commission's SME Performance Review 2016/17

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

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

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

Section A
Application Details

A1	Responsible Owner		
Full Name: [REDACTED]		Position: [REDACTED]	
Address: [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: [REDACTED]			

A2 Applicant Details (if applicant is not the responsible owner)

Full Name: [REDACTED]	Position
Address: [REDACTED]	Email: [REDACTED]
	Telephone: [REDACTED]
	Organisation: [REDACTED]

A3 Project Information

Project Title: Study of Small and Medium-sized Enterprises (SMEs) in the UK as part of the European Commission's SME Performance Review 2016/17

Start Date: 15/02/17

End Date: 30/07/17

Project Sponsor (select all that apply)

- ☐ ONS
☐ ADRN
☐ GSS
☐ Collaboration
☐ Other (Please specify).....EU Commission.....

A4 Collaboration and Sponsors

List of Collaborators/Sponsors	Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)
EU Commission	

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 (please specify).....

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)

The European Commission has commissioned London Economics, a policy and economic consultancy, to conduct an annual review of the performance of the SME sector and the national policy developments in the EU28 Member States and in Albania, Iceland, Former Yugoslav Republic of Macedonia, Moldova, Montenegro, Serbia and Turkey.

The SME Performance Review is one of the main tools the European Commission uses to monitor and assess countries' progress in implementing the [Small Business Act](#) (SBA) on a yearly basis. With an emphasis on the measures from the SBA Action Plan, the review brings comprehensive information on the performance of SMEs in EU countries and other

partner countries. It consists of two parts: an annual report on European SMEs and SBA country fact sheets.

Besides reviewing recent trends and the short-term outlook for SMEs, the 2016/17 report will also contain a special chapter focusing on self-employment, which will focus on employment creation by enterprises created originally as self-employed or one-person enterprises and will contain aggregate figures for net employment creation by firm type, sector and firm size

In order to track employment by enterprises created originally as self-employed or one-person enterprises and examine how the firms' employment performance varies across sectors, different firm characteristics and cohorts, London Economics have requested access to the de-identified ONS Business Structure Database (BSD) in the VML.

The study is focused on the contribution to employment of enterprises created originally as self-employed or one-person enterprises. This contribution will be compared and contrasted with the employment performance of other SMEs. The researchers will use the Business Structure Database to distinguish between enterprises created originally as sole traders, one-person enterprises and other SMEs. The SMEs will be classified and considered according to their size, and their date of creation.

The project is the first of its kind to relate employment generated in enterprises originally created as self-employed or as one-person enterprises to employment growth, and will have a key role in informing both UK and EU policy on entrepreneurship and self-employment.

Published annually, the SBA fact sheets aim to improve the understanding of recent trends and national policies affecting SMEs. Since 2011, each EU Member State has appointed a high-ranking government official as its national SME envoy. The UK SME Envoy is Susanna Simon, Director (Europe) at the Department for Business, Energy and Industrial Strategy (BEIS).

SME Envoys in EU countries are appointed by the national government. Their role is to

- promote SMEs' interests throughout all government bodies and ensure that the "[Think Small First](#)" principle is integrated into their policy-making and regulatory proposals;
- act as the main interface between the Commission and national policy-makers;
- report on the uptake of the SBA in EU countries, step up efforts to distribute information on SME policy actions, and promote the exchange of good practices.

London Economics prepared the 2015/16 Annual Report which showed that for 2015, EU28 SMEs as a group, experienced growth in value added, employment and the number of enterprises. In 2015, EU28 SME value added was well above its 2008 level (9 percent points) while SME employment still remained below (2 percent points). Besides reviewing recent and expected developments in SME employment and value added, and in the number of SMEs in the non-financial business sector, the report also examines in greater detail the SME employment creation dynamics, the SMEs' profitability and the SMEs' export performance in recent years in different Member States and economic sectors using both aggregate data from the Eurostat. In addition, the report examined the impact of the

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.

characteristics of bankruptcy regimes in the EU Member States on entrepreneurship
 Link to 2015/16 report > <http://londoneconomics.co.uk/blog/publication/annual-report-european-smes-201516/>
 Link to Small Business Act > https://ec.europa.eu/growth/smes/business-friendly-environment/small-business-act_en

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>				
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>			Business Structure Database	
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?
(max 500 words)

Access to de-identified BSD data will only take place within the secure ONS Virtual Microdata Laboratory (VML) environment and all outputs will be checked by the VML team prior to release to ensure disclosure control and the confidentiality of data subjects is protected. All analysis and use of the data will be within the VML at a secure VML setting at one of the ONS offices. The London Economic researchers have recently been accredited as ONS Approved Researchers following their successful completion of training on the Safe Use of Data in Research Environments (SURE) course. No data or drafts of the report will be shared with anyone outside of the project team and ONS will review the report before it is published. The researchers are very aware of the importance of protecting the confidentiality of the business subject analysed – particularly so given that many will be sole traders.

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

The research requires access to longitudinal micro enterprise data to track employment, net job creation or destruction by individual firms and examine how the firms' employment performance varies by cohort and firm characteristics. The study is focused on the contribution to employment of enterprises created originally as self-employed or one-person enterprises. This contribution will be compared and contrasted with the employment performance of other SMEs.

For this purpose, London Economics would like to gain access to the **Business Structure Database (BSD)** for their research. This data would enable them to distinguish between enterprises created originally as:

- sole traders;
- one-person enterprises; and
- other SMEs.

The dataset also allows for a distinction between firms that belong to larger entities (thus likely to have limited discretion in managerial decisions) and those that exist autonomously.

Once they have been able to separate these effects, for each year in which data is available, London Economics will distinguish three main groups of firms:

- soletraders created in that year;
- SME firms created in that year classified and considered by firm size; and
- SME firms which already existed in that year cassified and considered by firm size.

The researchers will then compare and contrast the employment creation of different cohorts

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

of firms starting out as self-employed or with only one employee across cohorts, types of firms and sectors.

The output of this research is therefore of a descriptive nature. There will be no regression analysis involved. A preliminary example of the data output is presented below:

Cohort 2009

	Sector	Employment/ Enterprises in year (t)	Employment/ Enterprises in year t + 1	Employment/ Enterprises in year t + 2
Sole traders created in 2009 (t)				
Firms created as sole traders which existed in 2009				
Firms created in 2009 (excluding sole traders)				
1-3 Employees				
4-9 Employees				
10-19 employees (if disclosure rules permit)				
20-49 Employees (if disclosure rules permit)				
50 – 249 employees (if disclosure rules permit)				
Firms which already existed in 2009 (Excluding sole traders)				
1-3 Employees				
4-9 Employees				
10-19 employees (if disclosure rules permit)				
20-49 Employees (if disclosure rules permit)				
50 – 249 employees				

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

(if disclosure rules permit)

Where no confidentiality issues arise, shares of these observations will also be computed, to identify each size class' contribution to the overall employment trends in a sector or in the overall economy.

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

The outcomes will provide an evidence base for identifying characteristics that have resulted in employment growth in businesses that were created as these types of enterprises and will be have a key role in informing both UK and EU policy on entrepreneurship and self-employment.

This will help further the understanding of the impact and economic consequences of self-employment on the UK and enable identification of the factors that inform policy and decision making in the UK.

The report allows the public to view the efforts of their own country in supporting SMEs, entrepreneurship and their level of compliance with the 10 principles of the Small Business Act whose purpose is to promote growth by creating the conditions for SMEs to flourish.

The 2015/16 SME performance review for the UK concluded that it has one of the most vibrant SME sectors in the entire EU. It considers that the UK has fully recovered from the financial crisis and continues to perform strongly. SME value added increased substantially during 2008 – 2015 and SME employment is estimated to have expanded by almost 7% during that period. The report evaluates the performance of the UK in implementing the Small Business Act for Europe (SBA) and concludes that the UK Government has continued its active business support strategy, led by the Small Business, Enterprise and Employment Act, 2015 and the 'Exporting is GREAT' initiative. This information will help the public hold the government to account about its performance and provide an evidence base to assess whether it is stimulating business growth through its policies and meeting its statutory requirements.

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

i. The use of data has clear benefits for users and serves the public good

The project will focus on the factors that influenced growth of enterprises who were sole traders or similar and will be used to inform UK and EU policy as well as annual review of the SME sector.

It will achieve a public benefit by providing an evidence base for decision making by policy makers at a national and international level and by SMEs looking to expand.

It will also allow the EU to include the UK in its annual report produced on SMEs.

The report will also help the public understand how the UK government is performing in its business growth policies and compare this with the performance of other EU countries. It will also help identify trends over the time. This will promote greater transparency and accountability

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 data will be accessed in the secure ONS Virtual Microdata Laboratory (VML) and is de-identified to protect the confidentiality of data subjects. The Approved Researcher scheme will be the legal gateway used to access the data and the researchers have successfully completed the Safe User of Research data Environments (SURE) training and have signed and will adhere to the Approved Researcher declaration setting out how they will manage the data and protect the confidentiality of data subjects in line with the Statistics and Registration Service Act 2007. The researchers will not be able to take the data outside of the VML and their outputs – both intermediate and final – will be checked and cleared by the ONS VML branch to ensure the confidentiality of data subjects is protected. ONS will see a copy of the final report prior to publication.

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.

No new technologies are being used. The research methods employed will be openly available for further scrutiny or replication of results.

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.

Access to the potentially disclosive data will be in a secure environment (VML) and via an approved legal gateway (Approved Researcher scheme). This is in compliance with the Statistics and Registration Service Act 2007. The methods used are compliant with the principles in the Data Protection Act.

v. The views of the public are considered in light of the data used and the

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

perceived benefits of the research.

Whilst the views of the public have not been sought with regards to the research, there is a clear public benefit for the analysis (as described in B6.1 above) and the outcomes from the research will be published in the 2016/17 Annual Report of European Small and Medium Enterprises (SMEs). A public consultation on the Approved Researcher scheme in 2015/16 recommended that commercial organisations should be allowed to access ONS research data where there is a clear public benefit and the ONS Microdata Release Panel (MRP) considered that this research will achieve a public good.

vi. The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public.

The research methodology and outcomes will be made public and reported on. London Economics aim to publish the research to inform and expand knowledge of how small firms have grown. The EU Commission will publish the findings as part of its annual SME report.

To help promote greater transparency and in compliance with the updated Approved Researcher scheme, the researchers have agreed to their details being included on a public record of Approved Researchers and to publishing the findings of their research (including on the ONS Approved Researcher pages). ONS will work with London Economics to consider the inclusion of this analysis as a published case study on the ONS website setting out the research methodology and outcomes, and their impact on public service delivery and decision/policy making.

B7 How will the findings of the research be disseminated?

The EU Commission will publish the findings as part of its annual SME report on its website. Previous years reports can be viewed on the EU Commission website by November 2017 at the following email address https://ec.europa.eu/growth/smes/business-friendly-environment/performance-review-2016_en#annual-report

The analysis will form the basis of a special chapter in the main SME annual report and a special appendix to the annual report based on all EU countries for which relevant micro data can be identified. The special chapter and appendix will describe the trends in net job creation by enterprises created originally as self-employed arising from the aggregation of micro data. This will be compared and contrasted by sector and cohort to trends in net job creation by SMEs more generally.

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.

--	--

None

1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10	
11.	

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:

London Economics would distinguish three main groups of firms: soletraders created in that year, SME firms created in that year split by firm size, SME firms which already existed in that year split by firm size.

Justification for focusing on these subsections or groups:

To meet the scope of the project by focusing on SMEs and sole traders

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

The ONS Approved Researcher scheme is the legal gateway being used to access the data. This is in compliance with the Statistics and Registration Services Act 2007.

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

N/A

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(17)12

Microdata Release Panel: Impact evaluation (Phase I and II) of the Innovate UK Technology and Innovation Centre network (Catapults)

Purpose

1. This paper presents a proposal referred by the Microdata Release Panel for access and use of ONS data by Belmana, an independent research and policy consultancy, to evaluate the performance of businesses that have received support from Innovate UK by studying business output, employment and productivity for the five years after the businesses received the support.

Recommendations

2. Members of NSDEC are invited to consider the 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 the proposal be resubmitted to a future meeting once implemented; or
 - iv. reject the proposal advising it be stopped from proceeding.

Background

3. Innovative UK is the UK's innovation agency and is an executive non-departmental public body, sponsored by the Department for Business, Energy and Industrial Strategy. Innovate UK has launched a network of sector-specific Technology and Innovation Centres called Catapults. Catapults are not-for-profit, independent physical centres which connect businesses with the UK research and academic communities. Each of the 11 Catapult Centres specialises in a different area of technology.
4. This proposal will see Belmana use ONS Survey data (Business Structure Database, Annual Responders Database, Community Innovation Survey, Business Enterprise Research and Development Survey and the Inter-Departmental Business Register) in the ONS Virtual Microdata Laboratory (VML).
5. The purpose of the research project is to evaluate the performance of businesses that have received support from Innovate UK by studying business output, employment and productivity for the five years after the businesses received the support. The performance of these businesses in these areas will be compared with a control group who have not received support.
6. The research outcomes will help evaluate the impact of the investment in innovation and UK productivity and develop an evidence base to assess the effectiveness of providing direct government support to UK strategic industries.

Petros Saravakos, NSDEC Secretariat, Central Policy Secretariat, UK Statistics Authority, 11 April 2017

List of Annexes

Annex A: Application: Impact evaluation (Phase I and II) of the Innovate UK Technology and Innovation Centre network (Catapults), Peter Stokes, Office for National Statistics, 11 April 2017

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.
----	----	----	----	----	----	----	----	----	----	-----

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

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

Section A
Application Details

A1	Responsible Owner		
Full Name: [REDACTED]		Position: [REDACTED]	
Address: [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: [REDACTED]			

A2 Applicant Details (if applicant is not the responsible owner)

Full Name [REDACTED]	Position: [REDACTED]
Address: [REDACTED]	Email: [REDACTED]
	Telephone: [REDACTED]
	Organisation: [REDACTED]

A3 Project Information

Project Title: Impact evaluation (Phase I and II) of the Innovate UK Technology and Innovation Centre network (Catapults)

Start Date: May 2017

End Date: March 2021

Project Sponsor (select all that apply)

- ☐ ONS
- ☐ ADRN
- ☐ GSS
- ☐ Collaboration
- ☒ Other (Please specify).....Innovate UK.....

A4 Collaboration and Sponsors

List of Collaborators/Sponsors	Details and relevant documentation relating to collaboration (you may attach copies of relevant documentation)
<p>Innovate UK - An executive non-departmental public body of the Department for Business, Energy and Industrial Strategy (BEIS)</p> <p>https://www.gov.uk/government/organisations/innovate-uk</p>	<p>See attached document: 'Data Information Sharing Agreement_SQW Belmana BMG_Feb 2017'</p> <p>This defines the data sharing relationship between the</p>

catapults, Innovate UK and Belmana. It was drafted by Innovate UK and is consistent with Department for Business, Energy and Industrial Strategy (BEIS) agreements as Innovate UK reports to BEIS.

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 (please specify).....

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*)

Innovate UK, the UK's innovation agency, launched sector-specific Catapult Centres following the Hauser Review in 2010. Hauser recommended that, to close the gap

between the research base and industry, government should invest in ‘translational infrastructure’ through a network of Technology and Innovation Centres (TICs) in areas where the UK had particular strengths. The centres were to “transform the UK’s capability for innovation in specific areas and help drive future economic growth”. Catapults are not-for-profit, independent physical centres which connect businesses with the UK research and academic communities. Each of the 11 Catapult Centres specialises in a different area of technology.

Innovate UK is evaluating all the Catapults, but has chosen to manage this research programme grouping the Catapults. The two Catapults in this project are in the Emerging and Enabling Technology grouping. Innovate UK has then contracted different organisations to undertake the evaluations for the groupings. The approach used for each group is informed by scoping research that Innovate UK undertook as the first part of the research programme. Innovate UK will ensure that there are consistent evaluation approaches applied for all the Catapults.

Belmana, a research and policy consultancy, has been commissioned by Innovate UK to carry out a full impact and economic evaluation of two of the eleven Catapults. This research will focus on Satellite Applications and Digital.

The scope of the project is an impact evaluation, as defined by HMT Magenta Book on evaluation, of Innovate UK’s Digital and Satellite Catapults focusing on activities funded by the Government. Impacts would be direct economic benefits for businesses (productivity increases) and indirect economic impacts, such as technology spillovers.

The research will study the effects on business output, employment and productivity for the five years following support and compare performance to businesses which did not receive support but are statistically comparable in other economic characteristics.

The Catapults’ activities include collaborative research, support for businesses, such as training/mentoring/networking for start-ups. Catapults also provide research facilities that may be unaffordable for small businesses. The research will focus on an evaluation of the two Catapults and not an assessment of the industry sectors.

Phase 1: It will be a multi-year project with Phase 1(in 2017) focusing on linking de-identified businesses (by ONS in the VML) benefitting from the Catapults to the IDBR and establishing the best approaches to estimating the economic impacts on turnover & employment using the de-identified Business Structural Database. There are several thousand businesses with which the catapults have had contact with and lists of these organisations have been securely transferred to the Belmana research team by the Catapult Centres. The team will transfer this to the Virtual Microdata Lab and analysis will be undertaken in the secure setting.

Phase 2: The initial analysis carried out in 2017, will be followed by an interim study in 2018. This will refine the 2017 work and analyse further data from 2017. It will also inform the more detailed analysis planned for 2010. Any additional data from 2017 will be applied for through the ONS Microdata Release Panel. Findings from this study will be made available to ONS in advance of publication on the InnovateUK website.

Phase 3: For 2020, the full analysis and final impact evaluation will adopt a similar methodology to phases 1 and 2, applying any lessons learned from these. By linking the client businesses, the researcher will compare the performance of the assisted and comparison groups tracking their business performance in the BSD (covering turnover, employment, and survival) before and after the assistance period, and use difference-in-

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

differences (DiD) analysis to provide an indication of the potential impact of the Catapults' assistance on growth and productivity performance (with productivity identified using a combination of turnover and employment data held in the BSD). The comparison group will be drawn from the same types of business as the supported group and considered at an aggregate level. The ONS Annual Respondents Database (ARDx) and Business Enterprise, Research and Development (BERD) data will also be used as they collect data on whether a business undertakes R&D.

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>				
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>			ONS Business Structure Database (BSD) Annual respondents Database (ARDx) ONS Business Enterprise, Research and Development (BERD) Inter-departmental Business Register (IDBR) - for linking purposes only	
Census Data <i>(please specify year, e.g. Census 2011 in the relevant options adjacent)</i>				
Other <i>(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)</i>			Business names of companies who have benefited from assistance and those used for comparison purposes – supplied by Belmana (under a data access agreement with owners (Digital Catapult and Satellite Catapult))	

The Business Structure Database (BSD) has a small number of variables, so the focus is on those: turnover, employment, industry, location and type of business. Belmana will also derive additional variables such as whether an industry that a business belongs to is considered high technology. There are a number of studies that have identified such derived economic variables.

B3 How will information be kept confidential and data kept secure?
(max 500 words)

Access to the de-identified data will only take place within the secure ONS Virtual Microdata Laboratory (VML) environment where there are strict controls in place to protect the confidentiality of data subjects.

There are robust processes in the VML, including output checking by the VML team prior to release to prevent data subject disclosure and controls over access to de-identified data that researchers work under, to manage these risks. There are no additional new risks due to access to the data from Belmana or this project.

All analysis and use of the data will be within the VML at a secure VML setting at one of the ONS offices. The data will be linked within the VML environment. The Belmana researchers are longstanding ONS Approved Researchers and have successfully completed their training on the Safe use of data in research environments (SURE) course. No data or report drafts will be shared with anyone outside of the project team and ONS will review the report before it is published.

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

The purpose of the research project is to evaluate those businesses who have received support from Innovate UK by studying the effects on business output, employment and productivity for the five years after support and comparing performance to businesses who did not receive support but are statistically comparable in other economic characteristics.

There are three project phases:

Phase 1, 2017: The focuses on: (1a) linking de-identified businesses benefitting from the Catapults to the IDBR; and (1b) on early impact measurement to explore whether the supported businesses perform differently to the benchmark provided by comparable but unsupported businesses.

The early impacts of government support are likely to be on business survival, employment or turnover and the evaluation will use the Business Structures Database.

- For 1a, ONS will use the IDBR Enterprise Reference variable (ENTREF) to match to Companies House numbers for a list of companies that the catapults have supported. The researchers do not require access to the IDBR and all data provided to the researcher will be de-identified by ONS.
- For 1b, the researchers will statistically analyse the businesses benefitting from Catapult support relative to comparable businesses (who did not receive support). The BSD variables used will be:
 - size (BSD variable employment),
 - age (calculated from BSD registration data),
 - turnover (calculated from BSD deflated appropriately)
 - industry (BSD primary SIC),

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

- Government region (BSD variable GOR).
- The research team will also construct variables, such as “Hi-tech”, using the BSD SIC.
- Belmana have a "contact" list for both of the catapults included in the study, and this is quite exhaustive. However, the Catapults have also scored the intensity of the interaction with each contact, so Belmana have a measure of the support given, ranging from a limited level of interaction to significant support. The focus will be the latter group, and Belmana will test the sensitivity of results to this score as appropriate.
- There is a distinct list of contacts for each catapult, from different contact management systems. They are similar in size. For the foreseeable, Belmana will analyse the two catapults separately rather than pool them as they are both large enough a sample and the two catapults have different logic to their support.
- The control group is identified through statistical matching. Propensity Score Matching (PSM) selects a business that looks like the supported businesses, in terms of observable characteristics prior to support. So, if the Digital Catapult tends to support small, but growing, London-based, IT firms, then the PSM would select similar businesses in the unsupported population. A second strategy is to use PSM, but on the pool of businesses that contacted the Catapult but were given a modest amount of support, using the interaction scores (level of support provided). The extent of the comparability can be tested, using so called balance and bounds tests. If the matching is sound, then any difference in difference on outcome measures after support can be attributed to the support. We will look at difference in difference on employment initially to see whether supported businesses grow faster than unsupported. However, the attribution will be more sophisticated with time, as more variables become available and as any impacts mature over time. The proposed technique to establish if the supported businesses perform differently to unsupported will be difference-in-difference.
- Methods will be refined as appropriate following peer review as InnovateUK and BEIS have an evaluation team that provides expert input as needed.

Phase 2, 2018: An interim study will be undertaken in 2018, refining the 2017 analysis and analysing the additional year of data (this data will be requested via the ONS Microdate Release Panel (MRP)).

Phase 3, 2020: This will involve full analysis and final impact evaluation drawing on the lessons learned from Phases 1 and 2 and additional years of survey data to help assess the effects of support on business output, employment and productivity over a five year period. As well as additional years of ONS data (Belmana will apply for any additional data through MRP), the project involves a survey of the client businesses, conducted by the company BMG Research under the same overall contract as this work for Innovate UK. BMG Research will not have access to the ONS survey data. Subject to approval to link the BMG research data to ONS survey data, matching of de-identified businesses will be requested to help compare the supported businesses with aggregated data from comparable businesses not receiving support.

As the final methodology and data requirements for Phase 3 will take account of the lessons learned from Phases 1 and 2, Belmana will submit a separate MRP/NSDEC project application for Phase 3 to clarify the research methodology and data requested. This NSDEC application requests the approval of Phases 1 and 2 only.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

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

The final research outcomes will help evaluate a major government investment in innovation and UK productivity using an approach that is deemed high quality by the Maryland Scale (quasi-experimental approach).

The research is being sponsored by a public authority (Innovate UK) and it is evaluating a publicly funded policy. So, decisions to procure the research have been informed by public benefit as Innovate UK is committed to making the evaluation publicly available. The research will achieve a public benefit by evaluating government policy on supporting innovation and providing greater public and parliamentary scrutiny and accountability on value for money of public investment, both in terms of the impact on sectors in which government is investing and in the ways the Catapults are investing. Business Secretary, Greg Clark recently announced an [independent review of Innovate UK's Catapult Centres](#). This research will contribute to the review.

The UK Government is committed to evaluating whether its policies deliver value for money. HM Treasury's Magenta Book and the Department for Business, Energy and Industrial Strategy annual evaluation plan make strategic commitments to deliver this and help to promote greater scrutiny and accountability. Innovate UK's 15/16 annual report states that "Measuring the impact of what we do is critical to enable understanding of the impact and value we create. Robust evaluation helps us to improve delivery and impact and, with more rigorous analysis of opportunities, will help steer future investments" (pg 20). This evaluation is part of a set of such research as major investments are made into the Catapults and was a recommendation of the Hauser Review. The findings will help the Government understand and ensure that it is targeting its support for businesses in the most effective way.

A further benefit from using the data available at the ONS VML is its potential to provide high quality evidence without a specific collection for the purpose of the evaluation. The use of public surveys – especially in looking at the performance of statistically similar but unsupported businesses – provides a comparison group so that additional impacts of support can be estimated without conducting a large survey of non-beneficiaries.

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*)

i. The use of data has clear benefits for users and serves the public good

This evaluation project will ensure an evidence base is available to analyse the merits of offering direct support to certain businesses and compare their performance to those not receiving the same support. Innovate UK is committed to understanding and quantifying the value for money of its publicly-funded support measures. The research outcomes will provide an evidence base for monitoring and evaluating policy delivery (support for UK businesses). It will support greater public scrutiny of government and promote greater transparency and accountability.

The evaluation is consistent with guidance set out in HM Treasury's [Magenta Book](#) and will be overseen by the Department for Business, Energy and Industrial Strategy (BEIS). A particular focus for BEIS is that the evaluation is consistent with other similar studies, including the evaluation of other Catapults. The use of the VML data has benefits in this regard, allowing common approaches to be used on a common dataset. The use also means a reduced burden on businesses as an alternative evaluation method would otherwise need data collections via a separate survey.

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 business data will be accessed in the secure ONS Virtual Microdata Laboratory (VML) and is de-identified to protect the confidentiality of data subjects. All matching will be carried out by ONS and the researchers will not be permitted to bring in any information to the VML or record any information from the data. The Approved Researcher scheme will be the legal gateway used to access the data and the researchers have successfully completed the Safe User of Research data Environments (SURE) training and have signed and will adhere to the Approved Researcher declaration setting out how they will manage the data and protect the confidentiality of data subjects in line with the Statistics and Registration Service Act 2007. The researchers will not be able to take the data outside of the VML and their outputs – both intermediate and final – will be checked and cleared by the ONS VML branch to ensure the confidentiality of data subjects is protected. ONS will see a copy of the final report prior to publication.

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.

No new technologies are being used. The research methods employed will be openly available for further scrutiny or replication of results.

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.

Access to the potentially disclosive data will be in a secure environment (VML) and will use an approved legal gateway (Approved Researcher scheme). This is in compliance

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

with the Statistics and Registration Service Act 2007. The methods used are compliant with the principles in the Data Protection Act.

v. The views of the public are considered in light of the data used and the perceived benefits of the research.

Whilst the views of the public have not been sought with regards to the research, there is a clear public benefit for the analysis (as described in B6.1 above). The research outcomes will be published on the Innovate UK website and this will support greater public scrutiny and evaluation of government policy and performance. It will help ensure that public money is spent efficiently and offers value for money. A public consultation on the Approved Researcher scheme in 2015/16 recommended that commercial organisations should be allowed to access ONS research data where there is a clear public benefit and it is the view of the ONS Microdata Release Panel that this work will achieve this.

vi. The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public.

The evaluation results will be published on the Innovate UK website. The results from Phase 2 study will be made available and published. The planned arrangements for the publication of the research findings from Phase 3 will be included in a separate (later) project application. Innovate UK already publishes the outcomes from its [evaluations](#). To help promote greater transparency and in compliance with the updated Approved Researcher scheme, the researchers have agreed to their details being included on a public record of Approved Researchers and to publishing the findings of their research (including on the ONS Approved Researcher pages). ONS will work with Innovate UK and Belmana to include this analysis as a published case study on the ONS website setting out the research methodology and outcomes, and their impact on public service delivery and decision/policy making.

B7 How will the findings of the research be disseminated?

Innovate UK has stated its approach to sharing evaluation reports. As part of their evaluation strategy, Innovate UK publish interim and final evaluation reports on the Innovate UK blog on the Gov.uk website which includes details of recent evaluations of the Catapults: <https://innovateuk.blog.gov.uk/2017/03/21/how-impact-evaluations-demonstrate-the-value-of-our-funding/>

This research project's outcomes will be shared through similar published reports as the stages of the research are concluded.

The main output will be a readable synthesis report, reflecting across the range of evidence about the impacts of the two Catapults (alongside the econometric work of this application, there is a qualitative evidence collection for case studies and a beneficiary survey that will be analysed). The more technical aspects will be annexed or in separate reports.

[These other two studies do not use ONS data. [SQW](#), an environmental consultancy, is

1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11.

leading a number of case studies and [Cambridge Econometrics](#) is looking at impacts at a sectoral level using a macroeconomic model. These studies are not within scope of this application.]

A link to the outcomes from the Belmana analysis and the overall aggregated and cleared results, will also be included on the ONS Approved researcher web pages in accordance with the Approved Researcher transparency criteria. ONS will have sight of the results before they are published.

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.

None

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:

Those businesses receiving direct support and comparable businesses who haven't received support will be the subject of the analysis. All data will be de-identified.

Justification for focusing on these subsections or groups:

The following groups will be selected within the scope of the research as described in section B4

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

The ONS Approved Researcher scheme is the legal gateway being used to access the data. This is in compliance with the Statistics and Registration Services Act 2007.

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

Belmana will provide a list of companies that the government has supported, which has been provided by the Catapults through a data access agreement solely for this study. The results from this will be analysed in the VML.

**Department for Education’s approach to data transparency and
privacy**

Oral presentation

Iain Bell

1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10	
11.	

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11.
----	----	----	----	----	----	----	----	----	----	-----

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11
----	----	----	----	----	----	----	----	----	----	----

Any other business

1.	2.	3.	4.	5.	6.	7.	8.	9.	10	11
----	----	----	----	----	----	----	----	----	----	----