Sixth Meeting of the
National Statistician's Data Ethics Advisory Committee

_____

Minute, Agenda and Papers

_____

Tuesday 11 October 2016

11:00 – 14:30

Board Room, UK Statistics Authority
Newport

National Statistician's Data Ethics Advisory Committee

Minute

Tuesday, 11 October 2016
Board Room, Drummond Gate, London

Present
Members
Mr Ian Cope (Chair)
Mr Robert Bumpstead (Deputy Chair)
Ms Vanessa Cuthill
Mr Keith Dugmore
Mr Colin Godbold
Mr Neil McIvor
Ms Isabel Nisbet
Ms Marion Oswald

UK Statistics Authority
Mr Adil Deedat
Dr Simon Whitworth

Office for National Statistics
Ms Meghan Elkin (for item 3)
Mr Richard Heys (for item 4)
Mr David Johnson (for item 7)

Welsh Government
Mr Martin Parry (for item 5 and 6)

Apologies:
Ms Annie Hitchman
Dr Dean Machin
Mr Osama Rahman
Professor Martin Severs
Mr Hetan Shah

1.   Minutes and matters arising from the previous meeting

1.1   The Chair welcomed members to the sixth meeting of the National Statistician's Data Ethics Advisory Committee (NSDEC). The Chair introduced two new members to the Committee; Vanessa Cuthill, Director of Research and Enterprise at the University of Essex and Keith Dugmore MBE, founder of the Demographics User Group.

1.2   The Chair informed members that the minute of the fifth meeting had been agreed by correspondence. The minute, agenda and papers from the last meeting are now published on the UK Statistics Authority website. One amendment from the minute was agreed.

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 since the last meeting the Chair had presented to a cross Government meeting of Departmental Directors of Analysis whilst the Deputy Chair presented to the Statistical Policy and Standards Committee. The Secretariat also presented a poster highlighting the service NSDEC offers at the Government Statistical Service Conference.

2.2   The meeting heard that NSDEC members Marion Oswald and Hetan Shah participated in a Radio Four programme about "the online identity crisis".

2.3   Marion also informed the committee of the recruitment of a PhD student to work on legal, ethical and privacy issues relating to the work of the ONS Big data team. Previous work undertaken by the team has included the use of twitter data to understand flow to work patterns.

2.4   The Chair provided the meeting with an update on projects previously considered by the committee. At the last meeting the British Broadcasting Corporation and the Office for National Statistics presented a proposal which would see ONS link Television Licensing data to the Census to further understand gender disparities identified as a part of the a recent independent review of licensing prosecutions. The meeting heard that recommendations made by the committee had now been implemented and the BBC and ONS are in the process of finalising details of the project.

2.5   The final draft of the safeguarding policy and guidance for staff has been agreed along with a form to record concerns. Training for interviewers is now being developed and will be delivered later this year in briefing the National Study of Health and Wellbeing: Children and young people, which will take place in January 2017. Interviewers have now been briefed on the safeguarding policy and will apply it across all surveys they work on. Feedback on the policy will be gathered after six months, along with that of the safeguarding contacts and the Chief Safeguarding Officer who will receive and consider concerns. Provision of training for safeguarding contacts and for senior staff who may undertake the role of Chief Safeguarding officer is being explored with the Social Care Institute for Excellence.

2.6   The information sheet and consent form for the para-data project has now been agreed. The Survey Methodology and Statistical Computing Division are now looking to implement this within their work.

2.7   The meeting heard that the Memorandum of Understanding between University College London and the Office for National Statistics was still being drafted and access arrangements for the tool are being considered. The Memorandum of Understanding and access arrangements will be shared with the committee once they are finalised.

2.8   The Chair updated the committee on Data Access Legislation which forms part of the Digital Economy Bill. The meeting heard that the bill was in committee stages with evidence being heard from Professor Sir Charlie Bean and Mr Hetan Shah amongst others.

3.   Administrative Data Census Outputs-Income

3.1   The Chair invited Ms Meghan Elkin, who is leading on this work within ONS's Census Transformation Programme, to present this item. The meeting heard that the Census Transformation Programme was exploring how DWP and HMRC data sources could be used in the production of income estimates.

3.2 Ms Elkin outlined the user need for income data and the public benefit such as to better understand deprivation to enable better targeting of resources and services.

3.3 The meeting heard that income questions had never been included in the Census as previous Census tests have shown it to cause a reduction in overall response rates. The recent consultation on the 2021 Census showed the public to be more willing for information relating to their income to be derived from administrative data rather than be included in a Census questionnaire.

3.4 Ms Elkin outlined existing methods of producing income statistics, the majority of which are derived from surveys. The meeting heard that ONS had access to some employment and benefits data; however there are some types of employment data such as self employment, and some benefits data which ONS have yet to access which could be integral to producing comprehensive income statistics.

3.5 The meeting heard that research is being undertaken within the Census Secure Research Environment where all data is anonymised before analysis. To further protect privacy, incomes above an agreed threshold have been aggregated to a common value in order to mitigate against the disclosure of individuals with high incomes.

3.6 It was suggested that care would be needed in communicating the initial income estimates produced. Stakeholders should be made aware of any statistical quality issues with the estimates.

3.7 Members suggested that as ONS acquire data for 2015 and 2016, consideration should be given to the changes in pension arrangements which could impact on the analysis.

3.8 Members were satisfied with the general use of the data for the purpose of producing aggregate de-identified statistics for publication later this year.

3.9 The meeting agreed that the Census Transformation Programme should complete an application form detailing their plans to use PAYE, benefits and other data sources to produce granular income statistics. Moving forward, uses of data sources relating to income will be subject to full ethical review.

Action: Census Transformation Programme to develop formal application on use of pay as you earn and benefits data to produce income estimates.

4. **Economic Statistics Centre of Excellence (ESCOE)**

4.1 The Chair invited Mr Richard Heys, Deputy Director within the Office of the Chief Economic Adviser in the Office for National Statistics, to present this item.

4.2 The meeting heard that the Bean review had recommended ONS, in conjunction with suitable partners from academia and the user community, establish a new centre of excellence for the analysis of emerging and future issues in measuring the modern economy.

4.3 My Heys informed members that ONS will fund ESCOE as an independent research centre, shaped by ONS's priorities for the development of economic statistics. A proposed model for engagement with NSDEC was also outlined which would see ONS engage with the committee when making data available to ESCOE.

4.4 The meeting heard that research specification will be agreed in December.

4.5 Members expressed some concerns that ESCOE appeared similar in nature to the Administrative Data Research Network (ADRN). It was suggested that ESCOE engage with relevant parties from the ADRN to learn from their experiences particularly around security accreditation and burdening data owners.

4.6   It was suggested that the successful tender have proposals reviewed by their own ethics committee before consideration by NSDEC.

Action: Richard Heys to meet with Professor David Hand, Chair of the ADRN Board to learn from the work of the ADRN.

5.   **Welsh Government: Exploring differences in the reporting of Welsh Language ability: An exploratory study linking survey and Census data. [NSDEC(16)11]**

5.1   The Chair declared an interest as information asset owner for the Census and passed chairing responsibilities to his deputy, Mr Bumpstead. Ms Nisbet also declared an interest as part of her role in Qualifications Wales.

5.2   Mr Bumpstead recapped NSDEC's role in considering research proposals from government researchers who require access to ADRN infrastructure.

5.3   Mr Bumpstead welcomed Mr Martin Parry, a senior Welsh Language officer from the Welsh Government to present this item. Mr Parry informed the committee that as part of the Well-being of Future Generations Act, a key goal was for a thriving Welsh language.

5.4   Mr Parry explained that there were differences between the number of Welsh speakers estimated from the Census and the National Survey of Wales (NSW). Both of which ask identical questions relating to Welsh Language. In order to understand this difference the proposal would look to link the NSW and the Census to understand the demographic groups who report their Welsh language ability differently on each of the sources.

5.5   Members suggested that the application should clarify, using non technical language, how confidentiality would be maintained. The length of time for which the data will be retained should also be included and time scales for the research should be revised.

5.6   NSDEC agreed that the project could proceed subject to these minor revisions being implemented.

Action: Secretariat to work with Mr Parry to implement recommendations before the proposal proceeds.

6.   **Welsh Government: An Exploratory study for estimating the outflow of Welsh Speakers from Wales to England from the 2011 Census and the Patient Register [NSDEC(16)12]**

6.1   Mr Parry introduced the item. The meeting heard that the aim of the proposal was to understand the outflow of Welsh speakers from Wales to England. Mr Parry explained that in the context of the Welsh Government trying to create a Wales with a thriving Welsh language, understanding the outflow of Welsh speakers was important.

6.2   Mr Parry outlined the proposed method by which the outflow of Welsh speakers could be estimated, by linking the 2011 Census to the Patient Register in England for years following the Census.

6.3   Members suggested that the application needed to outline more clearly what the public benefits would be from the proposal and how the findings would be used in evaluating Welsh Government policy. The research protocol should also be clarified.

6.4   Members agreed that the application should make clearer, using non technical language, how confidentiality would be maintained. The legal gateway outlined for accessing the patient register should also be clarified to provide assurance that the

relevant permissions have been received from the data owner to use the data for the purposes of this study.

6.5   The committee therefore recommended major revisions to the application, and sought recommendations to be implemented before a redrafted application is circulated to the committee for reconsideration.

Action: Mr Parry to redraft application, implementing committee recommendations for further consideration.

7.   Visit to the Data Science Campus

7.1   Mr David Johnson welcomed NSDEC to the data science campus. The meeting heard that the campus was up and running and taking on apprenticeships for a period of two years.

7.2   Mr Johnson informed the committee of the initial research themes for the campus which would be completed in collaboration with academic, industry and government partners and which would include

      i.   the modern economy;
      ii.   sustainability;
      iii.   the UK in a Global context;
      iv.   urban future; and
      v.   society.

7.2   Mr Johnson reported that he had been in the early stages of discussions with the secretariat about the development of ethical guidelines which could be incorporated into training that is being offered as part of the data science campus work programme. It was suggested that this would be an important piece of work and resource would have to be devoted to it to produce something suitable.

7.3   Mr Johnson reported that he has had discussions with various other departments such as the Department for Environment, Food and Rural Affairs, the Department for Culture, Media and Sport and the Department for Work and Pensions who are looking to develop some sort of ethical review process. It was reported that these departments are interested in hearing more about the work of NSDEC. Mr Johnson offered to put the secretariat in touch with relevant contacts in these departments.

Action: Secretariat to continue working with Mr Johnson to develop ethics training for data scientists.

8.   Any other business

8.1   The meeting heard that member Isabel Nisbet presented at a recent conference of government social research.

UK STATISTICS AUTHORITY

NATIONAL STATISTICIAN'S DATA ETHICS ADVISORY COMMITTEE

Agenda

Tuesday 11 October
Board Room, Newport
11:00am – 2:30pm

Chair: **Mr Ian Cope**

Apologies: **Ms Annie Hitchman**
Dr Dean Machin
Mr Osama Rahman
Professor Martin Severs
Mr Hetan Shah

(11:00am to 12:00pm)

1 11:00am	Minute and matters arising from the previous meeting (Page 8)	Mr Ian Cope
2 11:05am	Chair's report (Page 14)	Oral Report Mr Ian Cope
3 11:20am	Admin Data Census Outputs-Income (Page 15)	Oral Report Ms Megan Elkin

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

(12:30pm to 2:30pm)

4 12:30pm	Economic Statistics Centre of Excellence (ESCOE) (Page 16)	Oral report Mr Richard Heys
5 1:00pm	ADRC-Wales: Exploring differences in the reporting of Welsh language ability: An exploratory study linking survey and Census data (Page17)	NSDEC(16)11 Mr Martin Parry
6 1:30pm	ADRC-England: An exploratory study of estimating the outflow of Welsh speakers from Wales to England from the 2011 Census and the Patient Register (Page 31)	NSDEC(16)12 Mr Martin Parry
7 2:00pm	Visit to the data science campus (Page 32)	Oral report David Johnson
8 2:30pm	Any other business (Page 33)	

Next meeting: Tuesday 24 January, 2017

Location: One Drummond Gate, London

1. Minutes

2. Chair's report

3. Income

4. ESCOE

5. Welsh

6. Welsh

7. Data science

8. AOB

These minutes were agreed by correspondence and signed off by the Chair and are published along with the agenda and papers from the previous meeting on the [UK Statistics Authority Website](#).

National Statistician's Data Ethics Advisory Committee

Minute

Tuesday, 5 July 2016
Board Room, Drummond Gate, London

Present
Members
Mr Ian Cope (Chair)
Mr Robert Bumpstead (Deputy Chair)
Mr Colin Godbold
Ms Annie Hitchman
Ms Isabel Nisbet
Ms Marion Oswald
Professor Martin Severs

UK Statistics Authority
Mr Adil Deedat
Dr Simon Whitworth

Office for National Statistics
Mr Jon Wroth-Smith (for item 3 and item 7)
Mr David Johnson for item 5
Ms Jen Woolford for item 6
Mr Peter Stokes for item 8

Cabinet Office
Mr Oliver Buckley (for item 4)
Ms Madeleine Greenhalgh (for item 4)

British Broadcasting Corporation
Mr Mark Chipperfield (for item 7)

Apologies:
Dr Dean Machin
Mr Neil McIvor
Mr Osama Rahman
Mr Hetan Shah

1.   Minutes and matters arising from the previous meeting

1.1   The Chair welcomed members to the fifth meeting of the National Statistician's Data Ethics Advisory Committee (NSDEC).

1.2   The Chair informed the meeting that the minute of the fourth 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 Secretariat updated the meeting with progress on actions from the previous meetings. Most actions were complete or in progress and the secretariat updated the committee accordingly.

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 presented at the trust, risk, information and law conference at the University of Winchester in April. The presentation focused on the use of big data techniques and new sources in the production of National Statistics and the opportunities, ethics and governance challenges associated with these.

2.2   The Deputy Chair and Mr Deedat had presented to Heads of Profession for Statistics (HOPS) earlier this month. This included informing HOPS of the work of NSDEC and outlining examples of projects which had been considered.

2.3   The Chair presented at a recent conference on ethical issues in big data hosted by Cancer Research and Winton Capital. The presentation focussed on ethical issues in data linkage and the measures being taken by the UK Government Statistical Service (GSS) to ensure robust oversight of data linkage proposals. The Chair also used examples of data linkage projects considered by NSDEC to illustrate ethical oversight in practice.

2.4   The meeting heard that engagement across government about the work of NSDEC will continue over the summer. The Deputy Chair will be presenting to the GSS Statistical Policy and Standards Committee, while the Chair will be presenting to the Departmental Directors of Analysis.

2.5   The Chair provided the meeting with an update on projects previously considered by the committee. At the last meeting the Office for National Statistics presented a proposal to use paradata to select participants for follow up research to understand experiences of respondents. Members agreed that the proposal could proceed subject to a consent form and information sheet being provided to the committee for consideration. Since the last meeting members have commented on both of these via correspondence, and have approved subject to comments being implemented.

2.6   The Chair informed the meeting that revisions recommended to Welsh Government in relation to their proposal to share data with the University of Bath have been implemented. Members heard that the fair process notice and associated regulations governing the collection of data allow for the sharing of the information requested from the University of Bath. As far as privacy is concerned, the Welsh Government will also be anonymising the school details relating to pupil records and will now be providing the data at middle super output area rather than lower super output area. In addition information at Key Stage 4 will now be provided as aggregate indicators of attainment rather than the results of each individual subject. This will further minimise the risk of an individual being identified. Welsh Government also confirmed that the research proposal itself has now been deliberated on by the University of Bath's Ethics committee, with written confirmation approving the research now received by Welsh Government.

2.7   The meeting heard that the Scottish Government have begun disseminating their methodology for linking data. Engagement is taking place with partners in the Administrative Data Research Centre in Scotland. Scottish Government will also share their methods with the Economic and Social Research Council and other centres in the Administrative Data Research Network.

2.8　　Members heard that the recommendations made to an initial proposal relating a study of end of life care for children had been fed back to the University of Southampton. The University of Southampton are in the process of bidding for funding to undertake the work. Once this has been secured an application will be submitted with members initial recommendations implemented.

2.9　　The Chair updated members on progress with identifying a suitable access arrangement for the ethnicity from names tool. A range of software has been identified which would lock the tool to an individual's computer preventing onwards transfer and would be time stamped leaving the tool redundant after the allocated time has elapsed. Potential users of the tool would need to demonstrate a clear public benefit from its use before access is granted. Once these arrangements have been finalised and a memorandum of understanding has been agreed between UCL and ONS, these will be circulated to members for further consideration.

Action: Secretariat to circulate memorandum of understanding to Committee when available.

3.　　Data Sharing

3.1　　The Chair invited Mr Jon Wroth-Smith to present this item. Mr Wroth-Smith informed members that, following the last meeting, he had implemented comments made by the Committee relating to how the data shares were presented.

3.2　　The Committee were satisfied with the changes made to the data shares annex and suggested that greater information relating to the potential uses be included in future updates.

3.3　　Members considered whether NSDEC would in the future deliberate on proposals to access particular data sets. While most data being acquired by the office may not be contentious in nature, some data sets such as those relating to health or income may need additional consideration.

3.4　　The Chair clarified the role of NSDEC with regards to data shares. The meeting agreed that access proposals relating to health or income data should be considered. It was suggested that the acquisition of such data be subject to full ethical review at a future meeting.

3.5　　As ONS considers exploring acquisition of commercial data, NSDEC suggested that these data also be included within the annex at future meetings.

Action: Secretariat to work with the Administrative Data Division to bring an application relating to acquisition of income data

4.　　Public views on government use of data science

4.1　　The Chair welcomed Mr Oliver Buckley and Ms Madeleine Greenhalgh from the Government Digital Service (GDS) to the meeting.

4.2　　Mr Buckley introduced the work his team has been doing to improve data quality, literacy and infrastructure across government.

4.3　　Ms Greenhalgh then provided an overview of public engagement work undertaken by IPSOS Mori to understand public views on use of data science by government. The research was led by the Government Digital Service and part funded by Science-wise and the Office for National Statistics. It included four sets of focus groups in different locations across the country and online surveys.

4.4   The meeting heard that case studies were used to facilitate public understanding of data science. It was reported that there was generally a broad acceptance of the use of data science by government once risks and benefits have been explained.

4.5   The presentation fed back public views on a hypothetical case study from ONS, which set out how mobile phone data could be used in the production of population estimates. The meeting heard that there was general acceptance that ONS use of such data was relatively unconentious. Participants understood the public benefit of population estimates to government and most saw the advantage of using location data, once the method had been explained.

4.6   Some participants expressed concerns with using mobile phone location data in isolation and questioned whether modelling techniques were sophisticated enough to produce accurate statistics. The biggest concern however was around the delivery of future public services, which are informed by ONS statistics and the negative impact on individuals of using inaccurate data to change public services.

4.7   Mr Buckley then provided the committee with an overview of ongoing work, communicating with the public to establish a contract between the citizen and the state for the use of public data.

4.8   It was suggested that there will be breaches of data and that any future framework developed by GDS needs to explicitly recognise this.

4.9   Members suggested that NSDEC consider how the ethical principles within the GDS framework fit with NSDEC's own principles.

Action: Secretariat to compare and contrast NSDEC ethical principles with GDS ethical principles and report back to NSDEC.

5.   **Centre of excellence for data science**

5.1   Mr David Johnson from Digital Services Technology and Methodology Directorate in ONS outlined the office's plan for a data science campus. Mr Johnson informed the committee that the data infrastructure and recruitment for the campus were underway and would be finished later in the summer. Given the short timeline to delivery Mr Johnson appreciated the need to balance the work of the data science campus with existing ethical structures.

5.2   Mr Johnson informed the Committee of relevant recommendations from the Bean review around the use and capability of data science and the need for ONS to increase it skills, understanding and application of new methods and data in producing statistics.

5.3   The meeting heard that the campus will be partnered with a number of academic institutions including the Universities of Bath, Cardiff and Lancaster. ONS and the GSS will work with these partners to deliver an MSc programme and PhDs. By 2017 it is expected that the campus will have between 30 and 36 staff which will rise to 100 over a five year period. It will also look to re-skill existing ONS staff and offer apprenticeships to school leavers.

5.4   The campus will consider how other data could be incorporated in to the production of official statistics. For example data from Uber and Airbnb could be used in the production of economic statistics.

5.5   Early projects within the campus are likely to include research on educational attainment and its impact on longevity. This would make use of historic education data, graduate surveys and ONS data such as the Census.

5.6   Mr Johnson concluded by seeking advice from NSDEC on how the volume of data science projects could be accommodated by existing governance structures.

5.7 Members thanked Mr Johnson for a comprehensive presentation. It was suggested that information relating to those projects that are likely to begin before the next meeting will be circulated to members for consideration by correspondence.

5.8 The next meeting will be held in Newport, and members will be able to visit the data science campus and get more detailed perspective of the work.

Action: Secretariat to provide initial projects for the data science campus by correspondence should they begin before the next meeting.

6. **Draft safeguarding policy and guidance [NSDEC(16)09]**

6.1 Ms Jen Woolford from the Life Events division in ONS introduced the paper. Since the last meeting NSDEC comments had been implemented and the draft policy and guidance have been discussed at the GSS Statistical Policy and Standards Committee and at a meeting of the National Statistics Executive Group. It was reported that the policy and guidance would be piloted later this year.

6.2 Members welcomed the changes made to the draft policy and guidance. It was suggested that the policy and or guidance should include information on how staff would be supported following initial reporting of a safeguarding incident. Training should also be provided to staff.

6.3 Members suggested that incidents be rehearsed at different times, for example during out of office hours.

6.4 Members invited Ms Woolford to return a future meeting to update the committee on the pilot.

7. **Application: Television licensing prosecutions: gender disparity [NSDEC(16)10]**

7.1 The Chair welcomed Mr Jon Wroth-Smith from the Administrative Data Division in ONS and Mr Mark Chipperfield from the British Broadcasting Corporation to the meeting.

7.2 Mr Wroth-Smith introduced the item and explained that following an independent review of TV licensing a gender disparity had been identified, where considerably more women compared with men were prosecuted for not having a TV license.

7.3 Mr Chipperfield informed the meeting that the results of the analysis would provide more information on the groups impacted. This would help the BBC to better understand reasons for the disparity and consider whether current stakeholder activities are sufficient and whether training materials require updating.

7.4 Members sought clarity around consent and suggested the BBC further explore data sharing in line with their privacy statement. The meeting heard that only aggregate anonymised data would be shared with the BBC.

7.5 It was suggested that the BBC may wish to review the proposed contract with ONS and the BBC's third party. As it currently stands ONS would be a data processor; however members suggested that the BBC may want to show ONS as the controller.

7.6 Members agreed that the proposal could proceed subject to minor revisions.

Action: Secretariat to work with Administrative Data Division and British Broadcasting Corporation to revise the application.

8.   Changes to the Approved Researcher

8.1   Mr Peter Stokes from the ONS Public Policy Division introduced the item. The meeting heard that following consultation the Approved Researcher process had been changed. The changes will see researchers trained and accredited once while projects would still be considered on a case by case basis.

8.2   The changes to the Approved Researcher process would also see commercial organisations eligible to access ONS data in the Virtual Microdata Laboratory for the first time. Such organisations will however need to demonstrate a clear public benefit from their proposal to use the data.

8.3   It was suggested that NSDEC consider proposals for access from commercial bodies to ensure that they are ethical and demonstrate public benefit.

8.4   It was agreed that following a number of commercial projects being considered by the Committee precedents could be set which would facilitate deliberation on other proposals similar in nature.

Action: Secretariat to circulate link to Approved Researcher when published later this month researcher.

9.   NSDEC self-assessment

9.1   The Secretariat introduced the item and provided a summary of the outcome of the self assessment.

9.2   It was reported that the views of members were generally positive. There were suggestions for the inclusion on the Committee of an individual with a relevant social research background.

9.3   It was agreed that between meetings, the secretariat will provide members with an update. This will include:

   i.   relevant media coverage relating to ONS and the UK Statistics Authority;
   ii.   relevant publications;
   iii.   an update on relevant meetings and conferences attended by members; and
   iv.   an update on progress for projects implementing revisions.

   Action: Secretariat to provide updates between meetings

10.   Any other business

10.1   There was no other business.

Chair's report

Mr Ian Cope

1. Minutes

2. Chair's report

3. Income

4. ESCOE

5. Welsh

6. Welsh

7. Data science

8. AOB

Admin Data Census Outputs - Income[1]

Ms Meghan Elkin

---

[1]
https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/populationcharacteristics/researchoutputsincomefrompayeandbenefitsfortaxyearending2014

1. Minutes

2. Chair's report

3. Income

4. ESCOE

5. Welsh

6. Welsh

7. Data science

8. AOB

Economic Statistics Centre of Excellence

Oral report

Mr Richard Heys

1. Minutes
2. Chair's report
3. Income
4. ESCOE
5. Welsh
6. Welsh
7. Data science
8. AOB

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(16)11

Exploring the differences in reporting of Welsh Language ability: An exploratory study linking survey and Census data

Purpose

1. This paper presents a proposal from the Administrative Data Research Centre in Wales. The aim of the research is to explore differences in reporting of Welsh language ability at an individual level. The proposal would involve the linkage of survey responses from the National Survey for Wales with the 2011 Census. These sources have identical questions on Welsh language ability allowing for comparisons in reporting.

Recommendations

2. Members of NSDEC are invited to consider the project application at **Annex A** and advise the National Statistician to:

   i.   approve the proposal and allow it to proceed;
   ii.   approve the proposal subject to minor revisions;
   iii. recommend major revisions to the proposal and request the proposal be resubmitted to a future meeting once implemented; or
   iv.   reject the research advising that the research be stopped from proceeding.

Background

3. The Welsh Language is of great importance to Welsh Government. The Well-being of Future Generations (Wales) Act has as one of its well-being goals a "Wales of vibrant culture and thriving Welsh language". The Welsh Government also has an ambition to have one million Welsh speakers by 2050.

4. Ability in and usage of Welsh language is monitored through use of the National Survey of Wales and the Census, both of which ask identical questions on ability to speak Welsh. Comparisons between these sources show significant difference, with household survey estimates of Welsh speakers traditionally higher than Census estimates.

5. Through the Administrative Data Research Network (ADRN) the discrepancy between these sources can be assessed. Record level information would be linked between the National Survey for Wales and the Census to explore differences in reporting of Welsh language at an individual level, which may be due to differences between instruments, measures, or genuine differences over time.

6. This study will feed into the evidence base for policy making around Welsh language. It will also inform the work of the Census Transformation Programme whose work includes consideration of sources (other than Census) in measuring characteristics such as Welsh language.

7. The ADRN Approvals Panel, which assesses projects based on a number of factors including feasibility have approved the proposal subject to ethical approval.

Adil Deedat, Data Governance and Policy, UK Statistics Authority, 27 September 2016

List of Annexes

Annex A   **Application: Administrative Data Research Centre Wales- Exploring the differences in reporting of Welsh Language ability: An exploratory study linking survey and census data, Martin Parry, Senior Welsh Language Statistical Officer.**

National Statistician's Data Ethics Advisory Committee

Application for Ethical Review

1. Minutes

2. Chair's report

3. Income

4. ESCOE

5. Welsh

6. Welsh

7. Data science

8. AOB

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

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

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

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

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

Section A

Application Details

A1	Responsible Owner

Full Name: Martin Parry	**Position:** Senior Welsh Language Statistical Officer
Address: Knowledge and Analytical Services Welsh Government Cathays Park Cardiff CF10 3NQ	**Email:** ▮▮▮▮▮▮▮▮▮▮▮▮
	Telephone: ▮▮▮▮▮▮
	Organisation: Welsh Government

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: MARTIN PARRY

UK Statistics Authority

R. Martin Parry

Signature:
Date: 30/09/2016

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

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

A3	**Project Information**

Project Title: Exploring differences in the reporting of Welsh Language ability: An exploratory study linking survey and Census data (ADRN Project Number PROJ-0115)

Start Date: November 2016   **End Date:** January 2017

Project Sponsor (select all that apply)

☐ ONS

☐ ADRN

☐ GSS

☐ Collaboration

☑ Other   (Please specify) Welsh Government (Knowledge and Analytical Services)

A4	**Collaboration and Sponsors**

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

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 proportion and number of people speaking and using Welsh is of considerable importance to Welsh Government. One of the well-being goals of the Well-being of Future Generations (Wales) Act is "A Wales of vibrant culture and thriving Welsh language". The

1. Minutes
2. Chair's report
3. Income
4. ESCOE
5. Welsh
6. Welsh
7. Data science
8. AOB

new Welsh Government has set out an ambition to reach a million Welsh speakers by 2050. In order to measure the number of Welsh speakers, and in particular monitor the ambition of a million speakers by 2050, it is important to be clear about the robustness of the data sources and the coverage of Welsh speakers within those sources. This project will explore measurements of Welsh language ability in the Census and the National Survey for Wales (NSW), in order to examine differences in reporting Welsh language between these research instruments.

For the purposes of this study, a Welsh speaker is defined as an individual that has self-reported that they are able to speak Welsh in either the Census or the NSW.

Ability in, and the usage of, Welsh language is monitored through the use of both national surveys, such as the Welsh Language Use Survey, and at a population level through the Census. The Census is seen as the authoritative source for benchmarking the number of Welsh speakers. However comparisons between Welsh language ability between survey and Census data highlight significant differences.[1] Household survey estimates of the number of Welsh speakers are traditionally higher than that of Census estimates. This may be due to the differences between instruments, measures, sampling bias vs population data, or genuine differences over time.[1,2]

The ADRN provides the opportunity to link together - for the first time - survey data with Census data in order to explore differences in reporting Welsh language ability at an individual level. Specifically this study seeks to explore inconsistent reporting of Welsh language ability i.e. are individuals reporting Welsh language ability on the Census and not on surveys, and vice-versa? The NSW and Census both include identical questions asking whether an individual is able to speak Welsh or not (along with understanding spoken Welsh, reading and writing Welsh).[2]

This project will examine the demographic and regional differences in Welsh language ability, in order to identify any common factors that may lead to their inconsistent Welsh language ability reporting.

This project has implications for Welsh Government in gaining a more accurate picture of the Welsh language base in Wales. Furthermore, this project will also be of critical importance for the Census Transformation Programme in understanding the implications on data quality from a move away from a traditional Census. If in future there is a great reliance on surveys it will be fundamental to understand how those surveys compare with Census and the nature of the disparity between the Census and other sources, and what this means for change over time.

1. Haselden, L (2003) Differences in estimates of Welsh language skills. London: ONS
2. Bryson and Huskinson (2012) The National Survey for Wales: question comparability study. London: Bryson Purdon Social Research

B2	Data Use

Type of data	Data Level Please specify the name of the data set			
	Aggregate Data	Identifiable Data	De-identified personal data	Anonymised/ pseudo anonymised

Administrative data *(please specify, e.g. Patient Register 2011, School Census 2012 etc, in the relevant options adjacent)*				
Big Data *(please specify e.g. Twitter data, smart meters and mobile phones, in the relevant options adjacent)*				
Survey Data *(please specify e.g.LFS, BRES, etc in the relevant options adjacent)*			National Survey for Wales (2012-2013, 2013-14 and 2014-15)	
Census Data *(please specify year, e.g. Census 2011 in the relevant options adjacent)*			Census 2011	
Other *(please specify e.g. Ordinance Survey Address register in the relevant options adjacent)*				

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

The linked data set accessed by the researchers will only be de-identified data and the data linking will be done by a trusted third party so that staff and researchers cannot see information which would identify individuals, for a strict separation of activities: https://adrn.ac.uk/protecting-privacy/de-identified-data/trusted-third-parties

Data will be accessed from either the ADRC-Wales safe setting, or an ONS setting (dependent on the security requirements when linking to the Census): https://adrn.ac.uk/about/research-centre-wales

Persons accessing either safe setting will undergo training in the use of these settings, in addition to training on statistical disclosure. The training undertaken will be the ADRN accreditation training to become an 'accredited researcher': https://adrn.ac.uk/media/1162/adrn013-accreditedresearcher_v0100_pub.pdf

![UK Statistics Authority logo]

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

Data from the 2011 Census in England and Wales will be linked to data from the National Survey for Wales at the respondent-level. Matching will therefore be conducted on name, sex, age, address. The National Survey for Wales will form the spine point, to which will be appended Census responses. The final analytical dataset will therefore contain only information relating to individuals on the NSW and the Census – rather than the entire Census dataset. However, all individuals are requested from the NSW (even unlinked cases), in order to account for linkage biases and also as these individuals may have been absent during the Census period, yet present during survey years.

As the proposed study is exploratory, and the Census has not been linked to the National Survey previously, the methods used will predominantly be descriptive. The methods can however be grouped into several aims:

1. The first phase of the project will be to produce counts of people who say that they are able to speak Welsh in the Census but not in the National Survey for Wales, and those who say that they cannot speak Welsh in the Census but do in the National Survey for Wales. This will give a better idea of how both data sources compare in relation to people's ability to speak Welsh. Welsh-speaking ability will be defined as those that self-identified as being able to speak Welsh in both the Census and National Survey for Wales. This preliminary analysis will give us an insight into how individuals' responses may differ between the Census and the National Survey for Wales and will give us and stakeholders an indication of the reliability of household surveys to estimates the number and proportion of Welsh speakers.

2. The second phase of the project will be linking this information to wider information around Welsh language ability (i.e. reading and writing Welsh, fluency and frequency of speaking Welsh) and wider demographic information to understand the characteristics of those individuals that note different Welsh speaking ability on both sources. For example, we will be analysing the geographic location of these Welsh speakers to understand whether people in areas with high or low levels of Welsh speaking are most likely to report the same on both sources. This will be done by simple cross-tabs and significance tests. This will also be aided by producing maps to highlight any geographical trends.

3. Similar analysis will be done based on other demographic data such as age, gender, qualifications and economic activity. We will undertake logistic regression analysis to understand the relationship between these variables and the likelihood of reporting Welsh speaking "differently". This analysis will give us an insight into the characteristics of those individuals who report different Welsh language ability between the Census and the National Survey for Wales. For example, are younger people more likely to respond differently between the two surveys? Are people in the areas with the lowest concentrations of Welsh speakers more likely to respond differently between the two surveys?

Analytical note:
The National Survey for Wales asked questions of household members who were 16 or older at the time of interview, however as the Census was conducted in 2011, this means that individuals were potentially 14 years olds in the Census. The analysis will be dependent on the final population who report some form of Welsh language ability.

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

This project has huge benefits to the evidence base around the Welsh language. Increasing the numbers of Welsh speakers is an important priority for Welsh Government and they and public authorities are required to work towards the goal of a thriving Welsh language under the Well-being of Future Generations Act. As part of this they are required to undertake needs assessments and set well-being objectives which will be based on evidence. To understand fully the prevalence of Welsh language, particularly in years between Censuses, they need to interpret the difference in Census and survey data. This will help produce informed policy to support the promotion of the Welsh language.

The Census Transformation Programme is aimed at developing recommendations for modernising the Census in future by using more administrative data. This may lead to future Censuses being radically different and therefore an understanding of the different sources on the Welsh language is an important component for that project. A successful Census Transformation Programme has potentially huge benefits for society, value for money and the statistical infrastructure of the nation.

The key impact of the study will be to feed into the evidence base for policy-making around Welsh language. Improvements in the evidence base would be expected, in due course, to deliver societal benefits in terms of meeting the goal within the Well-being of Future Generations (Wales) Act 2015 of a vibrant culture and thriving Welsh language.

The findings of the research will be presented to the Welsh Language Policy Board in the Welsh Government and to other groups such as the Welsh Language Partnership Council, chaired by the Minister for Lifelong Learning and Welsh Language, who is responsible for giving advice and making representations to Ministers in relation to the Welsh language strategy. Policy officials and key stakeholders will be able to fully consider the outcomes and implications of the research study in light of the new Welsh language strategy, to be launched in March 2017.

We would publicise this research throughout the community of Welsh language researchers and stakeholders, alongside our usual statistical communities such as the Welsh Statistical Liaison Committee. We would also disseminate the findings to statisticians involved in household surveys and as described above the Census Transformation Programme team.

It is envisaged that the findings from the research project may form the basis of an ADRC Wales workshop/knowledge exchange seminar, in order to share the experience of using the Census for research purposes.

We would also seek to publicise the results through presentations for example in academic conferences or the National Eisteddfod for Wales, given the importance of the findings to the Welsh Government's cultural commitments.

UK Statistics Authority

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 the data has clear benefits for users and serves the public good. This project has substantial benefits to the evidence base around the Welsh language. Increasing the numbers of Welsh speakers is an important priority for Welsh Government and they and public authorities are required to work towards the goal of a thriving Welsh language under the Well-being of Future Generations Act. As part of this they are required to undertake needs assessments and set well-being objectives which will be based on evidence. To understand fully the prevalence of Welsh language, particularly in years between Censuses, they need to interpret the difference in Census and survey data. This will help produce informed policy to support the promotion of the Welsh language. Findings from this study have the potential to feed into policy decision-making with regards to the future of the Welsh language in Wales. Ultimately, this will have the potential to ensure the vibrancy of the Welsh language and Welsh culture for future generations.

The data subject's identity will be protected and the information will be kept confidential and secure. Data will only be linked for those individuals who have given their express consent for data linking from the National Survey for Wales (please see section C2). As mentioned in B3, the linked data set accessed by the researchers will only be de-identified data and the data linking will be done by a trusted third party so that staff and researchers cannot see information which would identify individuals. Data will only be accessed from either the ADRC-Wales safe setting, or an ONS setting and relevant researchers will have undertaken the ADRN accredited researcher training. This will ensure that we will be compliant with legal requirements around data, such as the Data Protection Act and the Statistics and Registration Service Act.

There will be sufficient human oversight so that the methods employed are consistent with recognised standards of integrity and quality. The ADRC will be involved in the output checking for the project.

A report will be published based on the project as a Government Statistical Service (GSS) publication on the Welsh Government website. The report will be accompanied by an accessible 'short report' or executive summary that will present a summary of the research findings and conclusions arising from the research project in an accessible manner for those who have an interest in the findings of the project but not the statistical detail, including policymakers and the general public.

The identity of individuals will be protected, for example any data that represents fewer than five individuals will be suppressed to minimise the risk of being able to identify an individual or an attribute about an individual. Only information about individuals who have given their express consent to link their answers from the National Survey for Wales will be included in the study. All of the information will be kept confidential and secure; only accredited researchers will be able to see the linked de-identified data.

B7	How will the findings of the research be disseminated?

A report will be published as a Government Statistical Service (GSS) publication on the

Welsh Government website. The report will be accompanied by an accessible 'short report' or executive summary that will present a summary of the research findings and conclusions arising from the research project in an accessible manner for those who have an interest in the findings of the project but not the statistical detail, including policymakers.

An important aspect of this work will be supporting the production of metadata and user guides in relation to the data used for this project. In line with the conditions of access, a short summary of the research will be produced for the ADS.

(See also response to B5 above).

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 dataset will not be available for use by other researchers. However the meta-data and the coding used in generating the initial data set will be retained for future use in order to re-create the data set, and perform other linkages between survey and Census data. The linked data will only be retained for a period of 12 months and be destroyed securely at that point.

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: Project excludes individuals aged 16 years old or younger at the time of interview as consent to link was not asked of these individuals.	
Justification for focusing on these subsections or groups: See above. Other than the exclusion of those younger than 16 years old, the sample for this study is from a national sample survey and is therefore representative of the	

population in Wales (though consent to link biases may occur).

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

Consent to link survey data to administrative data were asked in both applicable waves of the National Survey for Wales. The question asked was as follows:

There is ANOTHER way you can help us that would make your survey answers even more useful. The Welsh Government would like to be able to link your survey answers to information that the NHS and other public organisations hold about you. For example - your GP, hospital and education records.

Your information will only be used for research purposes and your personal details will be kept completely confidential. If you don't want us to link your data, that's absolutely fine, your survey answers will still be valuable without being linked to other records.

Please take a look at this card – this will explain what the Welsh Government are asking.

Please take your time to read this and ask me any questions you have.

Do you agree for us to link your survey answers with other information held about you? (Yes; No; Don't know; Refused)

Only the individuals who responded "Yes" to this question will be included in this study.

Analytical uses of the Census for statistical purposes are permissible – it is collected for statistical purposes and falls under Section 33 of the Data Protection Act 1998. As this project attempts to assess the quality of survey measures of Welsh language ability, as measured through the National Survey for Wales – however this has wider applicability to other surveys – this use falls under the permissible uses of the Census.

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

Census of Population data will be obtained from the Office for National Statistics. It will be obtained under Section 43 of Statistics and Registration Service Act 2007.

1. Minutes

2. Chair's report

3. Income

4. ESCOE

5. Welsh

6. Welsh

7. Data science

8. AOB

UK Statistics Authority

National Statistician's Data Ethics Advisory Committee

NSDEC(16)12

An exploratory study for estimating the outflow of Welsh speakers from Wales to England from the 2011 Census and the Patient Register

This project is undergoing major revisions and will be published in due course

Visit to the data science campus

Oral report

Mr David Johnson

1. Minutes

2. Chair's report

3. Income

4. ESCOE

5. Welsh

6. Welsh

7. Data science

8. AOB

Any other business