National Records of Scotland

Data Protection Impact Assessment (DPIA)
Administrative Data Mid-year Population and Household Estimates Project
Version 1.1

)

Document Control

Title	**DPIA – NRS Administrative Data Project**
Prepared by	**NRS: Head of Admin Data**
Approved by	**NRS: Head of Census, Statistics & Registration**
Date of approval	***8 Aug 2019***
Review frequency	**As required**
Next review date	

Status Control

Version	Date	Status	Reason for Amendment
0.1	16/05/2019	Draft	First Version – unable to edit
0.2	27/06/2019	Draft	Second Version – Modified template, changes information flows, GDPR considerations
1.0	07/8/2019	For IAO approval	Security and Privacy review and amendments
1.1	22/08/2019	Published	Proofed for publishing

Part 1: Data protection impact assessment screening questions

These questions are intended to help you decide whether a DPIA is necessary. Answering 'yes' to any of these questions is an indication that a DPIA would be a useful exercise. You can expand on your answers as the project develops if you need to. You can adapt these questions to align more closely to project you are assessing.

1. Will the project involve the collection of new information about individuals?

No. The project re-uses administrative data collected by NRS and other public bodies for research and statistical purposes only. Under the Legal gateway Census Act 1920 Section 2(1) and Section 5(1) and Lawful basis for processing: processing personal data are 1(c) and 1(e) of Article 6 of the General Data Protection Regulation (GDPR) and condition 2(j) of article 9 are met.

2. Will the project compel individuals to provide information about themselves?

No. The project re-uses administrative data collected by NRS and other public bodies for research and statistical purposes only. Under the Legal gateway Census Act 1920 Section 2(1) and Section 5(1) and Lawful basis for processing: processing personal data are 1(c) and 1(e) of Article 6 of the General Data Protection Regulation (GDPR) and condition 2(j) of article 9 are met.

3. Will information about individuals be disclosed to organisations or people who have not previously had routine access to the information?

No. The information will be retained for use by authorised NRS staff only. A de-identified version of the data will be sent to the NSS Safe Haven where only authorised NRS staff named on PBPP Application 1617-0195 will have access to the data.

4. Are you using information about individuals for a purpose it is not currently used for, or in a way it is not currently used?

No. The data is used by NRS for statistics and research purposes only. The statistics and research exemption applies under Article 89 of GDPR and s19 and Schedule 2, Part 6 of DPA18.

5. Does the project involve matching data or combing datasets from different sources?

Yes**.**

6. Does the project involve you using new technology that might be perceived as being privacy intrusive?

No. We are using standard statistical software packages such as SAS and R to analyse the data. The data is analysed on a secure server**.**

7. Will the project result in you making decisions or taking action against individuals in ways that can have a significant impact on them? Will you profile individuals on a large scale?

NRS will not use the data collected to make decisions about or take action that will impact on individuals. No individuals or households will be profiled.

8. Will you profile children or target marketing or online services at them?

No. No contact will be made with individuals.

9. Is the information about individuals of a kind particularly likely to raise privacy concerns or expectations? For example, special category data such as health records or criminal records, or other information that people would consider to be private.

Yes. – Although no special category data is explicitly processed, a marital status of “in a civil partnership” in Scotland currently indicates that the individual is in a same-sex relationship.

10. Will the project require you to contact individuals in ways that they may find intrusive?

No. No contact will be made with individuals.

11. Is the project collecting personal data from a source other than the individual without providing them with a privacy notice (‘invisible processing’)

Yes. The data is used by NRS for statistics and research purposes only. The statistics and research exemption applies under Article 89 of GDPR and s19 and Schedule 2, Part 6 of DPA18.

12. Is the project tracking individuals' location or behaviour?

No aspect of administrative data mid-year estimates operations involves tracking of individual's location or behaviour. Any outputs will be at aggregate levels at certain geographies, as such disclosure control methodologies will be used if the output may disclose a small number of individuals.

NRS maintains a record of answers to the screening questions in order to document that the decision on whether to carry out a DPIA was properly considered. If after completing the screening questions you decided a DPIA is not necessary you must send a record your answers to the NRS Data Protection mailbox. The NRS Information Governance Team will review answers, and where appropriate ask the NRS Privacy Group for their opinion.

Decision of Information Governance Team / Privacy Group

DPIA Required: Yes / ~~No~~	
Reason for decision: Required due to special category data processing, invisible processing and for PBPP application.	
Name: NRS: Census 2021 Deputy Data Protection Officer	Date: 7 Aug 2019

Part 2: Data protection impact assessment report

Use this report template to record the DPIA process and results. You can start to fill in details from the beginning of the project, after the screening questions have identified the need for a DPIA. The template follows the process that is used in the ICO code of practice. You can adapt the template to allow you to record additional information relevant to the DPIA you are conducting.

For further guidance please refer to the NRS DPIA Policy and Guidance

Step 1: Describe the project and identify the need for a DPIA
Explain what the project aims to achieve, what the benefits will be to NRS, to individuals and to other parties. You may find it helpful to link to other relevant documents related to the project, for example a project proposal or business case. It is important to include information about the benefits to be gained from the project in order to help balance any risk identified in the DPIA. This can help inform decisions on the level of risk to privacy that is acceptable, when balanced against the benefits or other justification for the project. Is there a benefit to the public? If a statutory duty exists provide details of this. Also summarise why the need for a DPIA was identified (this can draw on your answers to the screening questions) and identify the legal basis for processing.
Population and household estimates are key to the delivery of many public services. The size, age, gender and geographic distribution of the population are important statistics. This information also drives statistics about changes in the population and the factors driving these changes. These statistics have a wide range of uses. Central government, local government and the health sector use them for planning, resource allocation and managing the economy. They are also used by people such as actuaries for pricing pensions, market researchers and academics. Population estimates are fundamental to the distribution of billions of pounds across the United Kingdom, for example via the Barnett Formula[1]. Indeed around £73 billion in total was managed by the public sector in Scotland in 2017/18[2], including around £8 billion of Grant Aided Expenditure[3].

[1] Funding Devolved Government. CIPFA Briefing. November 2014
[2] https://www.gov.scot/publications/government-expenditure-revenue-scotland-2017-18/ Table 3
[3] https://www2.gov.scot/Topics/Statistics/18209/2019-20settlement

The Admin Data projects main objectives are:

- To produce administrative data population and households estimates at a range of non-disclosive geographies and compare these to the mid-year estimates currently produced by NRS.
- To enable quality assurance and to improve the accuracy of the 2021 Census estimates
- To make recommendations on the use of administrative data for future censuses.

The Admin Data project benefits would be:

- To research whether population and migration statistics can be produced to a higher level of quality or in a more timely fashion using individual record data from a range of sources. Improved population estimates between census periods will benefit the public through improved allocation of resources.
- To explore the benefits and limitations of administrative data as an alternative to a traditional census.
- To explore the costs of producing population estimates in this way. Current estimates are that the 2021 Census is expected to cost around £110m. Taking into account the limitations of administrative data, this will benefit the public by ensuring that the Census programme provides value for money by comparing alternative approaches.
- To research the viability of conducting future censuses using a combination of administrative data and surveys. This would have the benefit of reducing respondent burden.

The administrative mid-year population and household estimates will be compared with NRS's published estimates, so we will be seeking to produce administrative estimates at the following, non-disclosive geographies:

- Local Authority
- Locality
- Settlement
- Scottish Government Urban Rural classification
- Nomenclature of Units for Territorial Statistics (NUTS) - the statistical geography of the European Union,
- Scottish Index of Multiple Deprivation (SIMD) deciles,
- Scottish Parliamentary Constituencies (SPC)
- United Kingdom Parliamentary Constituencies (UKPC).
- National Parks

- Health and Social Care Partnerships (previously , pre April 2015, Community Health Partnerships)

The Admin Data Project in the UK context

Both the Office for National Statistics (ONS)[4] and the Northern Ireland Statistics and Research Agency (NISRA)[5] have made considerable progress in acquiring administrative data and developing new methods. ONS have now published research outputs, estimating the size of population in England and Wales for 2016 and 2017.[6]

Additional information

Any requests for changes to the data collection and methodology of this administrative project is submitted to the Public Benefit and Privacy Panel for Health Panel (PBPP) for approval. This Panel main aims are:

- Provides robust, transparent, consistent, appropriate and proportionate information governance scrutiny of these requests.
- Provides leadership and expertise for privacy, confidentiality, and information governance in relation to Health and Social Care in Scotland.
- Further strengthens the direct involvement of members of the public in the scrutiny process, and decision making regarding access to NHS Scotland originated data.[7]

[4] https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject
[5] https://www.nisra.gov.uk/statistics/2021-census/planning/reports-and-publications
[6] https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/sizeofthepopulation
[7] https://www.informationgovernance.scot.nhs.uk/pbpphsc/

Step 2: Describe the processing
Describe the nature of the processing: how will you collect, use, store and delete data? What is the source of the data? Will you be sharing data with anyone? You might find it useful to refer to a flow diagram or other way of describing data flows. What types of processing identified as likely high risk are involved?
In order to produce population estimates based on individuals, it's necessary to know the age and sex of each person and where they live, **but without knowing the identity of any particular individual or their exact location**. For household estimates it's necessary to know the number of people who live in a property, thus requiring a unique property reference number, rather than just the postcode. However, whilst personal information is required initially, on the advice of GCHQ (now the National Cyber Security Centre) this information will all be de-identified to preserve individual privacy. The potential for administrative data to replace information collected in the census will also be explored. Where available, ethnicity, disability and religion may also be collected to see whether the administrative data differs from that collected via the census. Where available, the date of last interaction with the service will also be collected. To begin with, the project proposes to just use the data held by National Records of Scotland in order to develop methods. These are as follows: • 2011 Census and 2011 Census Coverage Survey (reduced version with key variables only) • National Health Service Central Register • Vital Events – Births, Deaths, Marriages and Civil Partnerships • Scottish Address Database • Scottish Postcode Lookup Permissions have been granted and data sharing agreements are in place, for the following datasets: • ISD – Health Activity Dataset (this contains no health information, just the date of the last health interaction) • School Pupil Census • Higher Education Statistics Agency data on Scottish Students • Data on Scottish Further Education Students

- Electoral Register
- Census Coverage Survey for 2021
- Registers of Scotland (RoS) Residential Sales information.

The list of variables requested from each dataset has been reviewed and approved within the PBPP Application 1617-0195.

How will the information be collected?

The project involves using data gathered by the administrative systems of each data provider. No new data will be collected as part of the project. However, the project will use administrative data and data collected primarily as part of Scotland's Census 2021 – this includes 2019 Census Rehearsal Data, the 2021 Census and the 2021 Census Coverage Survey. The statistics and research exemption (Section 33(2) of the Data Protection Act 1998[8], superseded by Article 89[9] of the General Data Protection Regulations and section 19 and Schedule 2, Part 6 of The Data Protection Act 2018) allows for the re-use of administrative data for statistical purposes, provided the relevant conditions are met.

For each dataset, data will separate personal information – names, dates of birth, sex, addresses and postcodes - from the other information (known as the payload data). For the purposes of this study, payload data will include variables such as ethnicity, disability, religion and date of last interaction with service.

One member of the Admin Data team will access and separate the personal data from the payload data in a secure IT environment, once this process has been completed. Only the personal data with be transferred to another secure area and de-identification will occur here by a different member of the team. Only movement and deletion of data will have an audit trail that has been signed off by two members of senior staff. Table 1 explains how the various variables will be de-identified.

[8] http://www.legislation.gov.uk/ukpga/1998/29/section/33
[9] https://www.legislation.gov.uk/ukpga/2018/12/schedule/2/part/6/enacted?view=plain

Table 1: How the data will be de-identified.

Variable	De-Identified Derived Variable
First Name, Last Name, Date of Birth, Postcode	Used to produce a number of hashed[10] matchkeys which de-identify each individual.
Address and Unique Property Reference Number	Hashed Unique Property Reference Number. Maps each property to a de-identified one-way hash.
Postcode	Hashed Postcode, Hashed Postcode Sector Hashed Postcode District[11] Non-Disclosive Geographies
Date of Birth	Mapped to age in years on 30 June 2016 through to 2021, and 27 March 2021.
GP Practice Postcode	Using Eastings and Northings and Pythagoras's Theorem[12], the straight line distance between the GP Practice Postcode and the Patient's Postcode can be calculated. (For NHSCR and Health Activity Data only).

Who will have access to the data that is collected?

Access to the data will be by named individuals on Public Benefit and Privacy Panel Application (eDRIS 1617-0195). Limited members of Admin Data staff will have access to the personal identifiable information for the purposes of de-identification.

[10] Hashing is a one-way process which de-identifies the original data but retains its uniqueness. Hashing maps strings of different lengths to a sting of the same length.

[11] As geographies are likely to change over time the ability to future proof the geo-referencing will be needed. Indeed we'll need to include the 2021 Census geographies for the final evaluation of the work – This will be done by using hashed UPRNs and hashed postcodes to correctly assign properties to the appropriate geography, but maintaining the anonymity of each address.

[12] https://support.groundspeak.com/index.php?pg=kb.page&id=211

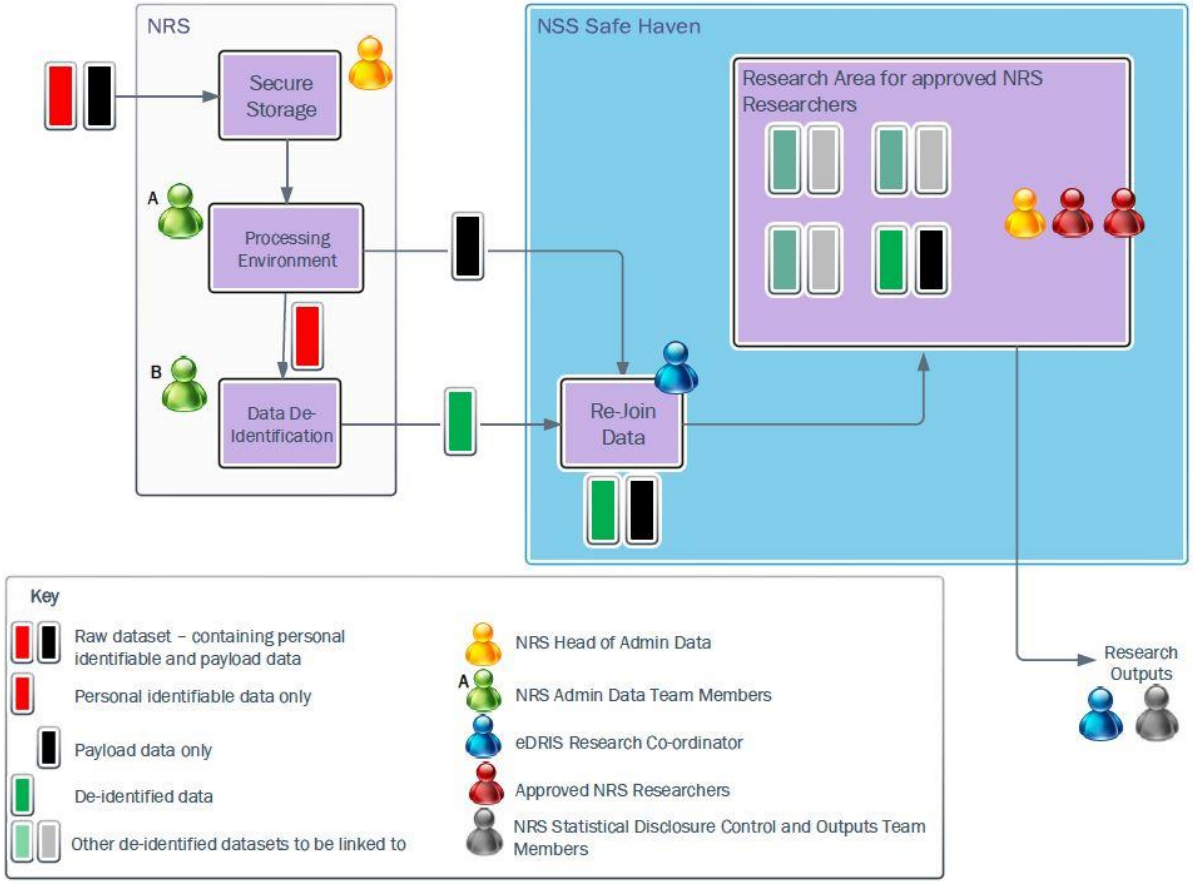
How will the data be collected and transmitted?

The Admin Data Project is using the Trusted Third Party model where there is a separation of functions across the projects. Separation comes from different team members within the team having distinct roles. One member of the Admin Data Team will be responsible separating the personal and payroll data and another for de-identification (the data de-identifier). These team member and the Head of Admin Data only will have access to the raw dataset.

1. Data providers send either just their personal identifiable data, or all of their data to the NRS Head of Admin Data team.
2. Head of Admin Data team stores each dataset on separate encrypted USB drives, stored in a fireproof box in a safe with an auditable lock. A backup safe with a copy of the datasets is kept in a different location.
3. The Head of Admin Data team transfers the dataset into secure location, (request is approved by senior manager).
4. Member of Admin Data team (person A) data splits each dataset into its personal identifiable data and payload data (the remaining variables), geographical data may be added at this time and added to the payload data. The payload data is encrypted using 7-zip.
5. Head of Admin Data team sent the Payload data separately to eDRIS co-ordinator at the NSS Safe Haven[13]
6. Head of Admin Data team requests personal data to be transfer to separate IT area for de-identification and deletion of personal data from person A security area (request is approved by senior manager).
7. Head of Admin Data transfers personal identifiable data to the data de-identifier (person B) who replaces the personal identifiable data with hashed matchkeys. This is done in isolation – one data source at a time. The resulting file is encrypted using 7-zip.
8. Head of Admin Data team sent de-identified data separately to eDRIS co-ordinator at the NSS Safe Haven. (Request is approved by senior manager).
9. Head of Admin oversees deletion of personal data from person B security area (request is approved by senior manager)
10. In the NSS Safe Haven, the eDRIS co-ordinator decrypts the hashed matchkey file and payload data and re-combines these files.
11. The file is then passed to the named NRS researchers (not person A or B), who combine all of the de-identified datasets to produce the statistical research outputs required for the project.

The eDRIS research co-ordinator performs statistical disclosure control on the outputs before they are released from the NSS Safe Haven. Where outputs contain 2011 Census data, 2011 Census Coverage Survey, 2019 Census Rehearsal data, 2021

Census data or 2021 Census Coverage Survey, these will also be checked by a member of Scotland's Census 2021 Statistical Disclosure Control team.



[13] http://www.isdscotland.org/Products-and-Services/EDRIS/Use-of-the-National-Safe-Haven/

How will personal information collected be stored, and disposed of when no longer needed?

Data will be stored in fireproof boxes in separate encrypted USB drives, in a safe with an auditable lock. Only the Head of Admin data and NRS Census Security personnel have access to the safe. A backup of the datasets is stored in the same way in a separate safe in a different location. At the end of the project, the data will be deleted from the drives in accordance with CPNI destruction standards and in accordance with the requirements of the NSS National Safe Haven.

Who will own and manage the data?

The data controller of the information supplied will be the Registrar General for Scotland. Day-to-day responsibility will rest with the Head of Admin Data team.

How will the data be checked for accuracy and kept up to date?

Once in the NSS Safe Haven, none of the data sources will contain names, dates of births or addresses. All the information in the Safe Haven will be de-identified. Named researchers will only have access to ages on specific dates and sex for each person in Scotland. Information on disabilities, ethnicity, religion and date of last interaction may also be known.

The various administrative data sources are likely to vary in quality. Through matching the de-identified data, we aim to work out the most likely de-identified address for each de-identified person in Scotland. This might be where two administrative data sources agree or, where there is a conflict across sources, we may use the person's most recent de-identified address as suggested by their most recent interaction.

Methodologies will be developed to improve the accuracy and quality of the statistics produced using the data that is collected. The result of the work will be compared with mid-year population and household estimates produced by NRS.

As the data is for statistical purposes, none of the work will have a direct impact on any individual

Describe the scope of the processing: what is the nature of the data, and does it include special category or criminal offence data? How much data will you be collecting and using? How often? How long will you keep it? How many individuals are affected? What geographical area does it cover?

Nature of the data

For each dataset, data providers will provide personal information – names, dates of birth, sex, addresses and postcodes. Some of the datasets will include variables such as ethnicity, disability, religion and date of last interaction with service. How we process this data has been explained in the previous section.

Although no special category data is explicitly processed, a marital status of “in a civil partnership” in Scotland currently indicates that the individual is in a same-sex relationship.

Data collection

The data collected from the above datasets will cover 2010/11 data to allow methodology checking to Census 2011. The other datasets to be received annually from 2016 to 2021. It is this aim the project to create administrative data mid-year estimates from 2016 to 2021. The project will hopefully be able to represent every person living in Scotland in order to create administrative mid- year estimates. The aggregated output from this de-identified dataset will cover the following non- disclosive geographies: Local Authority , Locality, Settlement, Scottish Government Urban Rural classification, Nomenclature of Units for Territorial Statistics (NUTS) - the statistical geography of the European Union, Scottish Index of Multiple Deprivation (SIMD) deciles, Scottish Parliamentary Constituencies (SPC), United Kingdom Parliamentary Constituencies (UKPC), National Parks and Health and Social Care Partnerships(formerly Community Health Partnerships).

Retention of data

The de-identified data transferred to the NSS Safe Haven may be retained for a period of five years after the end of the project in line with standard eDRIS safe haven procedures. The data held at NRS will be retained as per each individual data sharing agreement and shall be destroyed securely in accordance with CPNI destruction standards.

Describe the context of the processing: what is the nature of your relationship with the individuals? How much control will they have? Would they expect you to use their data in this way? Do they include children or other vulnerable groups? Are there prior concerns over this type of processing or security flaws? Is it novel in any way? What is the current state of technology in this area? Are there any current issues of public concern that you should factor in? Are you signed up to any approved code of conduct or certification scheme (once any have been approved)?
NRS will not be contacting individuals, the datasets are being used by NRS for statistics and research purposes only. The statistics and research exemption applies under Article 89 of GDPR and s19 and Schedule 2, Part 6 of DPA18. NRS are using a trusted third party ie NSS Safe Haven to create non-disclosive aggregate outputs. Every effort has been had to reduce persons having accessing to the unprocessed datasets. All processing of the personal data at NRS is auditable and held within a secure environment. All staff with the data Admin team have completed internal courses on Data Protection and Information Governance. The team have further completed the Medical Research Council online course on Information Governance, GDPR and confidentiality in order to use the NSS Safe Haven. All members of the team are part of the Government Statistical Service and as such are bound by the Code of Practise For Statistics[14].
Describe the purposes of the processing: what do you want to achieve? What is the intended effect on individuals? What are the benefits of the processing – for NRS, and more broadly?
Population and household estimates are key to the delivery of many public services. The size, age, sex and geographic distribution of the population are important statistics. This information also drives statistics about changes in the population and the factors driving these changes. These statistics have a wide range of uses. Central government, local government and the health sector use them for planning, resource allocation and managing the economy. They are also used by people such as actuaries for pricing pensions, market researchers and academics The benefits of this project:

[14] https://www.statisticsauthority.gov.uk/code-of-practice/

- To research whether population and migration statistics can be produced to a higher level of quality or in a more timely fashion using individual record data from a range of sources. Improved population estimates between census periods will benefit the public through improved allocation of resources.
- To explore the benefits and limitations of administrative data as an alternative to a traditional census.
- To explore the costs of producing population estimates in this way. Taking into account the limitations of administrative data, this will benefit the public by ensuring that the Census programme provides value for money by comparing alternative approaches.
- To research the viability of conducting future censuses using a combination of administrative data and surveys. This would have the benefit of reducing respondent burden

Step 3: Consultation process
Consider how to consult with relevant stakeholders: describe when and how you will seek individuals' views – or justify why it's not appropriate to do so. Who else do you need to involve within your organisation? Do you need to ask your processors to assist? Do you plan to consult information security experts, or any other experts? Describe the groups you will be consulting with and their interest in the project. Who should be consulted internally and externally? Explain the method you will use for consultation with any stakeholder groups and how you will communicate the outcomes of the DPIA back to them. How will you carry out the consultation? Explain what you learned from the consultation process and how they shaped your approach to the management of privacy risks. Explain what practical steps you will take to ensure that you identify and address privacy risks. You should link this to the relevant stages of your project management process. You can use consultation at any stage of the DPIA process.
NRS and the Admin Data Project is keenly aware of the wide diversity of interests represented by various groups and organisations who have a particular focus on what the census is and what it does. Engagement with our stakeholders will be key to helping us identify privacy risks and develop our plans to manage those risks. The Beyond 2011 project conducted a series of thirteen stakeholder events across Scotland during January and March 2013, and one public event in November 2012. The purpose of these events was to promote the Beyond 2011 Programme (data linkage of administrative datasets) and to seek feedback from a wide range of stakeholders. The Beyond 2011 team was superseded by the Scotland's Census 2021 Admin Data team in 2015 During 2017/18, the Admin Data team built on the previous stakeholder engagement and completed a lighter touch round of public and stakeholder engagement. This included meeting the Administrative Data Research Centre - Scotland Publics Panel, the ICO, privacy groups and local authorities The key points and suggestions from our stakeholders included: • Users require small area outputs, preferably down to datazone level. • User require a description of the mechanics of how the individual datasets are combined to produce population estimates. • There is the potential for alternative estimates of net migration. • The inclusion of enhanced outputs, particularly on income would be welcome. • Additional sources of information covering Council Tax, TV Licensing and Private Renting may be helpful to this project.

- We received positive feedback on how we have addressed privacy and security concerns.
- We were encouraged to hold public meetings – this was a comment from Open Rights Group during the consultation process in September 2017.
- The Data Privacy Impact Assessment (DPIA) should be published on our website and updated regularly.

The Public Benefit and Privacy Panel gave final approval for the project on 5th February 2018, assessing all conditions applied to the approval had been met. These conditions were that data sharing agreements were in place with all providers and to submit feedback from stakeholder and public engagement.

In 2018/19 we continued to deliver presentation and updates to our peers and stakeholders, though not as concentrated as in 2017. It is our aim that once we have created our first set outputs and after discussion with the Head of Profession for Statistics to release this as experimental statistics. Under the Code of Practice for Statistics, experimental statistics are published in order to involve users and stakeholders in the assessment of their suitability and quality at an early stage. At this point, there will be a high level of engagement with stakeholders, we will be targeting the main demography users. The information will be published on the NRS website to allow users who are not readily identifiable to contribute feedback.

Census Security and Privacy have reviewed this DPIA and have recommended its approval by the Information Asset Owner, the Director of Statistical Services. This is to be published as part of our stakeholder engagement.

Step 4: Assess necessity and proportionality
Describe compliance and proportionality measures, in particular: what is your lawful basis for processing? Does the processing actually achieve your purpose? Is there another way to achieve the same outcome? How will you prevent function creep? How will you ensure data quality and data minimisation? What information will you give individuals? How will you help to support their rights? What measures do you take to ensure processors comply? How do you safeguard any international transfers?
Under GDPR, we are processing this information: **Legal Gateway (also known as the power to share data)** The Parties are satisfied that the legal basis for sharing the Data comes from the Census Act 1920: Section 2(1) “Duty of Registrar-General to carry out census” It shall be the duty of the Statistics Board in relation to England and Wales and the Registrar General for Scotland in relation to Scotland to make such arrangements and do all such things as are necessary for the taking of a census in accordance with the provisions of this Act and of any Order in Council or regulations made thereunder, and for that purpose to make arrangements for the preparation and issue of the necessary forms and instructions and for the collection of the forms when filled up.” and section 5(1) “Preparation of statistics in respect of periods between one census and another” It shall be the duty of the Statistics Board in relation to England and Wales and the Registrar General for Scotland in relation to Scotland from time to time to collect and publish any available statistical information with respect to the number and condition of the population in the interval between one census and another, and otherwise to further the supply and provide for the better co-ordination of such information, and the Board or Registrar General for Scotland may make arrangements with any Government Department or local authority for the purpose of acquiring any materials or information necessary for the purpose aforesaid.” **Lawful basis for processing** The parties are satisfied that the lawful bases for processing personal data are 1(c) and 1(e) of Article 6 of the General Data Protection Regulation (GDPR):

“(c) processing is necessary for compliance with a legal obligation to which the controller is subject;”
“(e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;”

The parties are satisfied that for processing special category personal data condition 2(j) of Article 9 of the GDPR are met:
“(j) processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.”

The parties are satisfied that conditions 5(b) and 5(c) of Schedule 2 to the Data Protection Act 1998 (DPA) are met:
5 (b) the processing is necessary for the exercise of any functions conferred on any person by or under any enactment;
5 (c) for the exercise of any functions of the Crown, a Minister of the Crown or a government department;

The project has also using a third party Safe Haven model and this proposal have been approved by PBPP. The request to data providers has been limited to set a number of variables that would allow this project to proceed as per application proposal.
All the data requested so far are held by Scottish Organisations and all data processing will occur in Scotland.

All output from this project will be published as experimental statistics on the NRS website. Experimental statistics are developed under the guidance of the Head of Profession for Statistics and are published in order to involve users and stakeholders in the assessment of their suitability and quality at an early stage.

Step 5: Identify and assess risks				
Describe source of risk and nature of potential impact on individuals. Include associated compliance and corporate risks as necessary. Larger-scale DPIAs might record this information on a more formal risk register.				
No.	**Risk and potential impact**	**Likelihood of harm** (Remote, possible or probable)	**Severity of harm** (minimal, significant or severe)	**Overall risk** (low, medium or high)
1	**Unauthorised disclosure of information** There is a risk of compromise to confidentiality of information about individuals because of unauthorised or inadvertent disclosure leading to: • harm or distress to an individual or group of individuals; • reputational damage to NRS and/or the data providers; • possible enforcement action against NRS; • loss of confidence in NRS and/or the data providers; and, • loss of public finances to NRS and/or data providers. Examples include inadvertent compromise by a statistical process such as disclosure control, deliberate compromise by a member of staff or a targeted attack by cyber criminals.	Remote No evidence of release of census data accidentally or due to internal or external malicious activity.	**Significant (input data).** **Minimal (de-identified data)**	**Medium** **Low**
2	**Vulnerability in or malfunction of security controls** There is a risk of compromise to confidentiality, integrity or availability of information about individuals because of weaknesses in technical or procedural security controls leading to: harm or distress to an individual or group of individuals;	Possible Initial processing for de-identification will be in the	**Significant**	**High**

	reputational damage to NRS and/or the data provider; possible enforcement action against NRS; loss of confidence in NRS and/or the data provider; and, loss of public or non-public finances. Data subjects may have privacy concerns relating to the security of information technology used to process their data and about the extent of organisational measures in place to protect their data.	deprecated PARE (environment.)		
3	**Information is inappropriately shared between organisations .**	Remote	**Significant (input data)**	**Medium**

Step 6: Identify measures to reduce risk				
Identify additional measures you could take to reduce or eliminate risks identified as medium or high risk in step 5.				
No.	**Options to reduce or eliminate risk**	**Effect of risk** (eliminated, reduced or accepted)	**Residual risk** (low, medium or high)	**Measure approved** (yes, no)
1	**Unauthorised disclosure of information** • Datasets de-identified in isolation using hashed matchkeys. • Using Trusted Third Party Model, person doing de-identification is separate from researchers analysing the data. • EDRIS Research Coordinator and Census 2021 Statistical Disclosure Control team to check research outputs to ensure no-one can be identified from them.	Reduced	**Low**	**Yes**

	• All persons who may come into contact with census information will be required to sign the Census Confidentiality Undertaking which is underpinned by the Census Act 1920, prohibiting the sharing or unauthorised use of census data. This Act makes it a criminal offence, punishable by imprisonment, a fine or both, for any person to disclose any personal census information to another person without lawful authority. • NRS is committed to ensuring that privacy of every individual whose data will collected and processed as part of this programme will be protected. All statistical outputs produced by NRS fully comply with the Code of Practice for Official Statistics: Under section T6 on Data governance: Organisations should look after people's information securely and manage data in ways that are consistent with relevant legislation and serve the public good • A member of Admin Data team will de-identify personal identifiable data used in the study to minimise the amount of personal identifiable data used in the study. • NSS Safe Haven - The transfer of data into the NSS Safe Haven is controlled by the eDRIS research coordinator. Outputs are checked prior to release from the safe haven to prevent the disclosure of individual's privacy.			
2	**Vulnerability in or malfunction of security controls** • PARE server vulnerabilities are now actively managed although some structural issues (such as weak encryption) remain. • Identified datasets are not stored within PARE unless being actively processed.	Reduced	**Medium**	**Yes**

	• Datasets are stored on encrypted USB drives in fireproof box in a safe with an auditable lock. There is a separate drive for each data source. • Datasets are only moved onto the network when they are needed, in isolation and only for the time necessary for data processing. • De-identified datasets are processed on the managed and secure National Safe Haven.			
3	**Inappropriate sharing of input data.** • Input data will not be shared outwith the NRS Admin Data team. • Statistical disclosure control will be applied to de-identified data outputs. • NRS will continue to treat de-identified data as personal data and subject it to appropriate controls.	Reduced	**Low**	**Yes**

Step 7: Sign off and record outcomes		
Item	**Name/date**	**Notes**
Measures approved by:	NRS: Head of Census, Statistics & Registration	Integrate actions back into project plan, with date and responsibility for completion
Residual risks approved by:	NRS: Head of Census, Statistics & Registration	If accepting any residual high risk, consult the ICO before going ahead
DPO advice provided:	NRS: Census 2021 Deputy Data Protection Officer	DPO should advise on compliance, step 6 measures and whether processing can proceed
Summary of DPO advice: The Statistical Futures element of COP is not expected to have SAS processing capability until 7 Sep 2019. The current state of PARE server management is acceptable for the limited period of processing for the de-identification element of this project, with data subsequently being purged from the environment.		
DPO advice accepted or overruled by:		If overruled, you must explain your reasons
Comments:		
Consultation responses reviewed by:		If your decision departs from individuals' views, you must explain your reasons
Comments:		
This DPIA will kept under review by:	NRS: Head of Admin Data	The DPO should also review ongoing compliance with DPIA

Part 3: Linking the DPIA to the GDPR data protection principles

Answering these questions during the DPIA process will help you to identify where there is a risk that the project will fail to comply with the GDPR or other relevant privacy legislation, including the Human Rights Act.

GDPR Principle (a) (Article 5(1) (a))

Personal data shall be processed fairly and lawfully and, in particular, shall not be processed unless:

a) at least one of the conditions in Article 6 is met, and

b) in the case of special category personal data, at least one of the conditions in Article 9 is also met.

Have you identified the purpose of the project?

To provide create aggregate administrative data population and household estimates statistics, to make recommendations for future censuses in Scotland, with regards to:

- Whether administrative data can be used instead of a census to produce population and household estimates,
- The extent to which gaps in coverage and over-coverage in administrative data can be compensated for, and;
- The extent to which questions asked by a Census can be answered by administrative data

How will you tell individuals about the use of their personal data?

This will be done by the NRS Privacy Notice and publication of this DPIA.

Do you need to amend your privacy notices?

NRS continuously reviews its Privacy Notices to ensure that they reflect the current position. NRS will amend its privacy notices to indicate that NRS receives data from other public bodies and uses this for statistical and research purpose.

Have you established which conditions for processing apply?

The parties are satisfied that conditions 1(c) and 1(e) of Article 6 of the General Data Protection Regulations (GDPR) are met: (c) processing is necessary for compliance with a legal obligation to which the controller is subject; (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; The parties are satisfied that conditions 2(j) of Article 9 of the GDPR are met: (j) processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

If you are relying on consent to process personal data, how will this be collected and what will you do if it is withheld or withdrawn?

Consent is not being sought.

If your organisation is subject to the Human Rights Act, you also need to consider:
Will your actions interfere with the right to privacy under Article 8?

The data is already collected by the Scottish Government and Scottish departments and agencies. Processing of personal data by professional statisticians to produce aggregate statistics is not envisaged to present any additional interference with the privacy rights of individuals.

Have you identified the social need and aims of the project?

The use of accurate population statistics guides significant Government expenditure and provides material benefit to society.

Are your actions a proportionate response to the social need?

Yes. The approach identified enables the use of administrative data to meet the social needs and aims of the project, whilst balancing the need for individual privacy

GDPR Principle (b) (Article 5(1) (b))

Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.

Does your project plan cover all of the purposes for processing personal data?

Safeguards are in place to ensure that the data collected are only used for lawful purposes.

Have you identified potential new purposes as the scope of the project expands?

This will be considered as the project progresses. Any potential new purposes will need the approval of the Public Benefit and Privacy Panel

GDPR Principle (c) (Article 5(1) (c))

Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed.

Is the quality of the information good enough for the purposes it is used?

Extensive statistical methodologies and quality assurance processes will be put in place to ensure that the statistics produced using the information collected are fit for purpose and best meet the needs of data users. By definition, the Admin Data Project is of a research nature. It will look into whether administrative data can provide information of sufficient quality to produce population and household estimates.

Which personal data could you not use, without compromising the needs of the project?

The nature of the project dictates that we need to know basic demographic information about people – their age and gender- and where they live. Household estimates require that we know the address of individuals. We have taken reasonable measures to ensure that personal identifiable information is only gathered for the purpose of creating matchkeys.

GDPR Principle (d) (Article 5(1) (d))– accurate, kept up to date, deletion

Personal data shall be accurate and, where necessary, kept up to date.

If you are procuring new software does it allow you to amend data when necessary?

No. This is a statistical project. We'll be using SAS in the NSS Safe Haven. We will develop methodologies to deal with situations where administrative data sources contain conflicting information .

How are you ensuring that personal data obtained from individuals or other organisations is accurate?

Using a variety of administrative data sources, we will be able to quality assure the information supplied by data providers to check for conflicts and ensure accuracy of our final estimates**.** We are in the process of creating Quality Assurance of Administrative Data (QAADs) reports for each source, which will be published alone side our final estimates.

GDPR Principle (e) (Article 5(1)(e))

Personal data processed for any purpose or purposes shall not be kept for longer than necessary for that purpose or those purposes.

What retention periods are suitable for the personal data you will be processing?

Data will be retained for a period of five years after the end of the project in line with standard NSS Safe haven procedures, unless data providers have stated a particular retention period in their data sharing agreement.

Are you procuring software that will allow you to delete information in line with your retention periods?

No specialist software is being procured as part of the Admin Data project.

GDPR Articles 12-22

Personal data shall be processed in accordance with the rights of data subjects under this Act.

Will the systems you are putting in place allow you to respond to subject access requests more easily?

No. The data is used by NRS for statistics and research purposes only. The statistics and research exemption applies under Article 89 of GDPR and s19 and Schedule 2, Part 6 of DPA18.

If the project involves marketing, have you got a procedure for individuals to opt out of their information being used for that purpose?

Not applicable

GDPR Principle (f) (Article 5 (1) (f))

Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.

Do any new systems provide protection against the security risks you have identified?

.

No new systems are being procured for this project.

What training and instructions are necessary to ensure that staff know how to operate a new system securely?

No new systems are being procured for this project.

GDPR Article 24

Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country of territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.

Will the project require you to transfer data outside of the European Economic Area (EEA)?

No. All data will remain within Scotland

If you will be making transfers, how will you ensure that the data is adequately protected?

Data transfers between data providers, NRS and the NSS Safe Haven will all be via approved secure file transfer methods. This includes use of the Thru file transfer system or via secure FTP clients (such as Serv-U). Files will be encrypted using 7-zip prior to file transit to ensure files are secure during all stages of the transfer process.