

Report of a breach of the Code of Practice for Statistics

Core Information

Information needed	Response
Title and link to statistical output	The People and Nature Survey for England: Adult survey https://www.gov.uk/government/statistics/the-people-and-nature-survey-for-england-adult-data-y1q1-april-june-2020-experimental-statistics
Name of producer organisation	Natural England
Name and contact details of person dealing with report	Simon Doxford (simon.doxford@naturalengland.org.uk)
Name and contact details of Head of Profession for Statistics or Lead Official	Department for Environment, Food and Rural Affairs (Defra) Head of Profession – Ken Roy (ken.roy@naturalengland.org.uk)
Link to published statement about the breach (if relevant)	N/A
Date of breach report	24/11/2020

Circumstances of breach

Information needed	Response
Relevant principle(s) and practice(s)	Trustworthiness: Principle T6 (Practices T6.1 and T6.3)
Date of occurrence of breach	30/09/2020

Natural England's People and Nature Survey started data collection in April 2020. Natural England published the first release from the new survey of adult respondent data on 30 September 2020. This dataset contains variables on survey responses, demographics and also geographical data relating to the home location of the respondent and the green and natural places they visit. Natural England had commissioned a contractor – Kantar – to run the People and Nature survey on our behalf. Prior to the survey being conducted the parties agreed in writing that, for the purposes of the Data Protection Act 2018, Natural England is the data controller and Kantar is the data processor of the personal data collected in the survey and subsequently used by Natural England.

The full postcode data collected in the survey is used to accurately append important contextual geographical information for the respondent (Index of Multiple Deprivation Rank, Lower Super Output Area and Upper Tier Local Authority) – this process is carried out by Natural England. The full postcode is then converted to postcode sector (average 3000 households). Postcode sector is published in the People and Nature Survey dataset, thereby satisfying our obligation to publish the data under an open government licence on GOV.UK.

Contained within the published dataset were four coordinate fields derived from the full post code for 5119 respondents that were published in error by Natural England:

- Postcode latitude
- Postcode longitude
- Postcode easting
- Postcode northing

Natural England believed the four coordinate fields supplied were calculated from the postcode sector. However, these fields represented the centre point of the area covered by the respondent's full postcode (on average 15 households). There is a small risk that this information could be triangulated with demographic information provided in the dataset to identify a respondent's residence. This demographic information contained in the dataset includes:

- age (as a continuous and banded variable, 5085 respondents answered this question)
- gender (male, female, in another way (specify), all respondents answered this question)
- ethnicity (18 ethnic groups, 5091 respondents answered this question)

We believe this is a small risk due to the specialised technical skills required to process the data – which would involve plotting the coordinates and converting them back to the full postcode using Geographical Information Software – the number of respondents affected (see above), the short duration the data was available and requirement for pre-existing knowledge.

If a user were able to identify the residence of a respondent, this would contradict the privacy statement set out on the questionnaire “With your consent Natural England would like Kantar to collect and pass on your postcode to them to assign geographic areas such as Local Authority to your responses. Once this has been completed, Natural England will destroy the information.”

The quality assurance process for the People and Nature Survey was based on the lessons learned from a [data breach that occurred in 2017](#) as part of the Monitor of Engagement with the Natural Environment (MENE) survey – the predecessor to the People and Nature Survey – where full postcode data was published. Natural England was solely responsible for the 2017 data breach, which resulted from one of its employees uploading the wrong files onto Natural England's publication catalogue. The contractor for the MENE survey in 2017 was TNS, which was subsequently taken over by Kantar and whose team was largely disbanded. The Kantar team working on the latest People and Nature Survey are new to Natural England and had not previously worked on MENE. Natural England continues to receive full postcode information as part of the People and Nature Survey in order to join other geographic information on to the data file, for this reason our internal quality assurance (QA) process includes a specific check for postcode data in alphanumeric format. Full postcode was removed from the final dataset received from Kantar before publication.

Once the full postcodes were removed from the dataset, Natural England believed that the dataset was ready for publication and compliant with the Data Protection Legislation. However, the additional four geographical coordinate fields derived from the full postcode were not picked up through Natural England's quality assurance process, and were therefore published in error. The quality assurance process conducts a number of checks on each field including expected format, completeness and also data sense checks looking for results within an expected range and comparing results between similar questions.

This release was the first data-set under a new contract for a new survey. The breach has arisen due to confusion and differences in interpretation relating to the agreed data specification and the contractual requirements to deliver data-sets to a publishable standard. As all outputs from the research are intended for the public domain, Natural England believed the four coordinate fields supplied were calculated from the postcode sector and the final dataset provided by the contractor was ready to publish and fully compliant with legislation. Natural England require data that they publish to exclude specific geographic variables that could allow users to identify any respondent. Kantar provided a dataset that was fully compliant with legislation if the dataset were for internal use only by Natural England, but there was a misunderstanding between the two parties in relation to the requirement to provide a dataset ready for the public domain.

The breach was first identified on 21 October 2020 at 10am by Natural England. Steps were taken to immediately take down the data from the release page on GOV.UK and this was completed at 11.30am the same day.

Impact of the breach

The dataset was published on the 30 September 2020 and, in error, contained full post code coordinate information for 5,119 respondents.

In the intervening period the data was downloaded on a total of 32 occasions. A data download was the only method of viewing the data. As the data was hosted on GOV.UK and available publicly we are unable to identify the data users and contact these users to further mitigate the breach.

The risk of a respondent's residence being identified from the People and Nature Survey data published on 30 September 2020 is low due to the size of a typical postcode area. We have identified a risk arising from the potential combination of converted coordinate data alongside demographic information (e.g. age, gender and ethnicity). However, we believe that this risk is very small given the specialised technical knowledge required to process the data using GIS software, the number of data processing steps that would be required, the degree of local knowledge that would be needed to identify a respondent's residence, the requirement for the respondent to have a unique age, gender and/or ethnicity profile for that postcode, and the span of time the data was available for download.

After speaking with the Natural England Data Protection Manager and the Defra Data Protection Officer, it was confirmed that this breach does not meet the threshold for reporting to the Information Commissioners Office. This is because the published data relates to the centre of a postcode area (which covers 15 households on average) and does not directly identify a specific individual. Using this data also requires multiple data processing steps to derive a respondent's location. In addition there is the requirement for local / pre-existing knowledge in order to potentially identify an individual respondent using other information published in the survey dataset – or for the respondent to have a unique combination of these characteristics.

Corrective actions (taken or planned) to prevent re-occurrence

Once the error had been identified, the data was immediately taken down from GOV.UK and discussions began with Kantar to ascertain how this breach took place and actions needed to rectify it.

An internal data security incident report was filed, which triggered a subsequent internal investigation. This took account of lessons learned and corrective actions identified across the Kantar and Natural England teams.

The four coordinate columns mentioned above have been removed (we are not intending to add them back in with correct data) and the new dataset was been uploaded to GOV.UK on 30 October 2020.

We have used existing relationships to identify the users responsible for two of the data downloads and these users have deleted the relevant data from their datasets. We have alerted all of our known users – 66 users who are part of our Research User Group – to the data breach and requested that they delete relevant data and use the revised dataset going forwards.

We are now amending the Statement of Works via a contract change to make clear that all reports, data-tables and data-sets need to be delivered in a final format ready to be published on gov.uk consistent with relevant legislation. The contractual amendments will also strengthen arrangements around data provision to ensure data is provided to Natural England with appropriate security classifications. These classifications will be used to clearly mark datasets that still contain full postcode data in any format ('Internal official sensitive'), as well as datasets that do not contain

full postcode data in any format and are ready for the public domain ('public'). We have also worked with Kantar to review and identify derived products from full postcode data contained in previous datasets and have marked these accordingly ('internal processing only') on our internal systems.

The People and Nature survey interim indicators went through an [Office for Statistics Regulation rapid review in June 2020](#) and comments on our approach to quality assurance were as follows: 'The quality assurance process is proportionate and robust. Your team works closely with the contractor that runs the survey to check and validate the data. External stakeholders peer reviewed the bulletin, which allowed you to gather feedback on the indicators themselves as well as suggestions for improving your quality assurance process.'

We have further reviewed Natural England's quality assurance process and put extra measures in place to check no fields included in datasets for publication directly or indirectly reveal a full postcode. This will be added as standard to our QA process for the People and Nature Survey ([detailed here](#)) and we will ensure Kantar updates their own QA procedure to this effect. Natural England's Lead Official for Statistics will conduct additional final QA on special category data prior to final sign off. This issue, and all subsequent steps and corrective actions, will be detailed in the publication of our next technical report in December 2020. A note will also be added to the relevant GOV.UK page informing users to remove any adult survey respondent data downloaded before 21 October 2020.

We are also exploring the possibility of releasing datasets to users on a registration only basis in order to better handle sensitive fields and improve our ability to track users.