Table of content

Task 1:

1.1

Data Collection methods in the 2021 Census (England & Wales)
3 major techniques were deployed for data collection, which are:

- **Online Submission:** The 2021 Census was the first digital census; people were encouraged to fill out the form online using a unique share code sent to every household (ONS, 2025)
- **Paper Questionnaires:** Hard copies of census questionnaire papers were sent to some areas upon request, especially areas without internet access or preferring hard copies.
- **Field Visit:** 20,000 field staff were trained for non-response follow-up.

These 3 methods helped the ONS to achieve a 97% response rate after the adjustment.

1.1b

Yes, the ONS published a dedicated Privacy statement stating: "Your Privacy is important to us, your personal information will be kept safe and secure, ONS can only use this information for research" (Office for National Statistics, 2021). The ONS assured to follow the privacy policy in the Data Protection Act 1998 and the Statistics and Registration Service Act 2007.

1.2

Data storage is the process of keeping records organized, secure, and accessible only to authorized users. The ONS collects and processes Census 2021 data, storing it in secure, encrypted systems, ensuring it is not identifiable and accessed by researchers in a safe and secure form (ONS, 2025).
The UK Data Service provides the largest collection of data in the United Kingdom, offering long-term access and digital preservation. They work closely with the ONS to ensure the storage and accessibility of data. To access the most sensitive data, the researcher must be trained and gain Accredited Researcher Status administered by the ONS, and these data are only accessible via a secure remote environment and cannot be downloaded (UK Data Service, 2024).
The UK Data Service uses rich **metadata standards**, such as the **Data Documentation Initiative (DDI).** These records include information about methods, data file type, and access conditions, which together demonstrate the data's provenance.
Provenance is the detailed documentation of a dataset's collection history. To understand a dataset's authenticity, quality, credibility, and trustworthiness, providence documentation is essential. It is also mandatory for data to be Reusable under **FAIR** principles**: Findable, Accessible, Interoperable, Reusable** (GO-FAIR, no date). Researchers are limited in their analysis and interpretation without such documentation.

1.2b

Yes, not all stages of these collections were well documented and properly recorded, not giving a clear view of data transformation. For example, the process in which paper questionnaires were converted to digital form with optical character recognition (OCR) does not have enough clarity (ONS, 2023).

1.3

- **Duplicate removal:** The shift to online census form submission in 2021 led to a high rise in duplicate records due to online and paper submission from the same household. The ONS introduced the Resolved Multiple Responses (RMR) method to detect and merge these duplicates into single, coherent records(Office for National Statistics, 2022).
- **Imputation:** The ONS utilised the Canadian Census Edit and Imputation System (CANCEIS) for editing and imputation, copying donor values from similar responding households based on age, sex, and location(Office for National Statistics, 2023). They ensured coverage accuracy by using the Census Coverage Survey (CCS) (Office for National Statistics, 2022). The ONS compared records from the main census and the CCS records, using the Dual System Estimation (ONS, 2023). This helped them resolve under coverage and over coverage.
- **Consistency Check**: For consistency check the ONS verified row, column totals and value ranges (Office for National Statistics, 2022). This method not only maintains accuracy but also builds public confidence and trust towards census output.

1.3b

Data cleaning methods can correct inconsistencies and missing records, but can also introduce potential bias if the system fails.

- **Duplicate Removal**: There is a potential bias in the duplicate removal if is not critically looked into. For example, if two different people look alike such as twins sharing the same name or address, it can lead to undercounting.
- **Imputation**: filling in missing or inconsistent values using CANCEIS rises a potential bias. if imputation models make assumptions that don't fit all groups, like age, ethnicity, or region, they can affect the accuracy of the data.
- **consistency check:** strict rules may incorrectly flag or remove valid but unusual data. For example, a large extended family living together might be flagged as overcrowded when following the household vs room criteria.

1.4

- **Missing Data:** There is a possible **coverage error.** This is an error that occurs from failing to obtain some or all of the information from a member of the population, or when an answer to a question is missing, invalid, or inconsistent with the rest of the completed questionnaire (ONS, 2023). This error result to missing data, affecting the accuracy and quality of the census results.

- **Misreporting:** An error that can lead to misreporting is **measurement error** this occurs when respondent fails to fill in the correct information (ONS, 2023), as a result of misunderstanding what is required, responding multiple times or at the wrong address, all these can cause misreporting.
- **Classification drift:** One of the gradual changes that occurred in the 2021 UK census introduced a new "Roma" ethnic category and a write-in option under "Black African" to better reflect population diversity (Ethnicity Facts and Figures, 2021). These adjustments improve inclusivity and relevance and also complicate direct comparisons between census years(ChatGPT, 2025)

How it was mitigated?

Missing data were reduced through a digital-first approach that required filling of required fields and targeted follow-ups in low-response areas. **Misreporting** was addressed through online form validation and CANCIS edit rules(Office for National Statistics, 2022). **Classification drift**: The ONS maintains consistency of the new categories and make sure to publish the documentation that makes the census years similar (ONS, 2023). This approach is effective and has set a strong standard for future censuses.

1.5

It is the duty and responsibility of data professionals to ensure the confidentiality and privacy of the census data. They must ensure that census data are anonymised to avoid individuals from being identified when published**.** Data professionals must also maintain trustworthiness and credibility of data by ensuring accuracy, consistency, and reliability. They must ensure that census data are not used for discrimination or marginalization but are analysed and reported in a way that promotes equality across all communities. Whatever action taken data, professionals must ensure that they align with established ethical frameworks, they should be guided by the **ACM Code of Ethics, Virtue Ethics, and RSS Guideline.**

Task 2:

2.1

- **Sampling Bias:** The CCS achieved only 61% response against a 90% target (Office for National Statistics, 2022). As a Result, some groups, like the homeless, transient workers, remain undercounted. Such under coverage can deform resource allocation and present a failure of professional conduct under the **ACM code (1.2, 1.4, 2.5)**
- **Measurement Bias**: This occurs when a respondent fails to provide the correct information because they do not understand what is required (ONS, 2023). The ONS made their platform user-friendly; therefore, it passes the **ACM code 2.1**
- **Processing Bias**: Errors such as geographical assignment, data capture, editing, and coverage can occur during data processing before producing the final estimates (ONS, 2023). This bias should be avoided in line with the **ACM code of ethics (1.2, 1.3, 2.5)** to avoid harm and ensure fairness in the census results.

2.2

- **Consent:** Respondent were not informed that their geographic attribute may be modified. The ONS protects the confidentiality of individuals' data in statistical outputs using a technique known as record swapping, ensuring that no individual can be traced in published datasets (ONS, 2025). While the method protects privacy and meets legal requirements, but raises ethical concerns about consent and transparency. According to the **ACM Code 1.6**, computing professionals must respect individuals' privacy and ensure data use aligns with informed consent.
- **Fairness:** fairness means everyone is treated equally. While respondents online are being authenticated and validated to ensure their records are correct, what about the elderly respondents who filled in their data using the paper form have less error control? According to **ACM code 1.4**, all data collection method ensures equal accuracy and reliability.
- **Accountability:** Accountability means the public can review and question decisions. But come to think of it, the ONS uses the CANCEIS algorithm to edit data, and this algorithm is not publicly available. For instance, no one can verify whether the data recorded as a grandmother living with her granddaughter, "grandmother and grand daughter relationship" might be altered to "mother and daughter relationship". This raises concerns about data integrity and public trust. According to **ACM Code 1.3 and 2.5,** professionals should ensure transparency and openness to maintain accountability in data processing.

2.3

- **Application of ACM Code of Ethics (2018)**

  The ACM code **1.2 Avoid Harm** states that technology companies and professionals must ensure they avoid harm and that their work benefits society (ACM, 2018). In the UK Census 2021, the ONS used "CANCEIS donor-based imputation" to fill missing data (ONS, 2025, pp. 163-164). This method does not fill the actual missing but copies a record from a similar "donor" household to avoid statistical error.

If donors come mainly from majority groups, minority households such as multigenerational or single-parent households may be misrepresented, creating systematic bias violating ACM 1.2 by causing harm through inaccurate representation.

- **Application of Royal Statistical Society (RSS) Guidelines (2019)**

  The ONS adhered to **RSS Guidelines (2019), principle 4** (Building Trust Through Transparent Communication with the Public) by openly admitting the CCS shortfall, naming the four Local authorities requiring adjustments, and publishing the Alternative Household Estimates (AHE) methodology used to correct the estimate (Office for National Statistics, 2022). This demonstrates strong transparency in acknowledging and addressing limitations**.**

- **Application of Virtue Ethics (Professional Character and Responsibility)**

  Despite the underperformance of the Census Coverage, the ONS introduced the Alternative Household Estimate (AHE), along with NHS, Ministry of Defence, and school census data to correct undercounts in four Local authorities (Office for National Statistics, 2022). This demonstrates professional responsibility by going above and beyond basic census duties to ensure accuracy.

Comparison of Three Ethical Frameworks

- The ACM Code (2018) is rule-based and strict: it judges ONS imputation as failing due to bias risk in minority representation (1.2 Avoid Harm).
- RSS Guidelines (2019) are pragmatic: they acknowledged CCS limitations (61%) but credit transparent fixes (AHE, admin data). ONS meets standards.
- Virtue Ethics focuses on character: ONS earns strong praise for honesty, responsibility, and courage in admitting and correcting errors.

Conclusion: ACM is rigid, RSS is practical, Virtue is personal. RSS best fits real-world census work.

Task 3:

3.1

Data quality means completeness, accuracy, consistency, reliability, and freedom from bias. The ONS (2022) reports that CCS achieved only 61% against the 90% target, resulting in under coverage. Such incomplete data creates underfitting for population forecast AI models trained with the dataset.

Provenance is the traceability of data adjustments. The ONS (2022) used mixed-effects logistic regression to model response probabilities and applied bias corrections via the Alternative Household Estimate (AHE) and administrative data sources like the NHS, schools in five local authorities. Provence makes the system more reliable.

3.2

When census data is biased, AI systems for healthcare resource allocation that use the dataset for training will underperform or make inaccurate decisions. For example, low CCS participation and imputation gaps underrepresent ethnic minorities in inner-city areas, leading to underestimated health needs (ONS, 2022).

Task 4

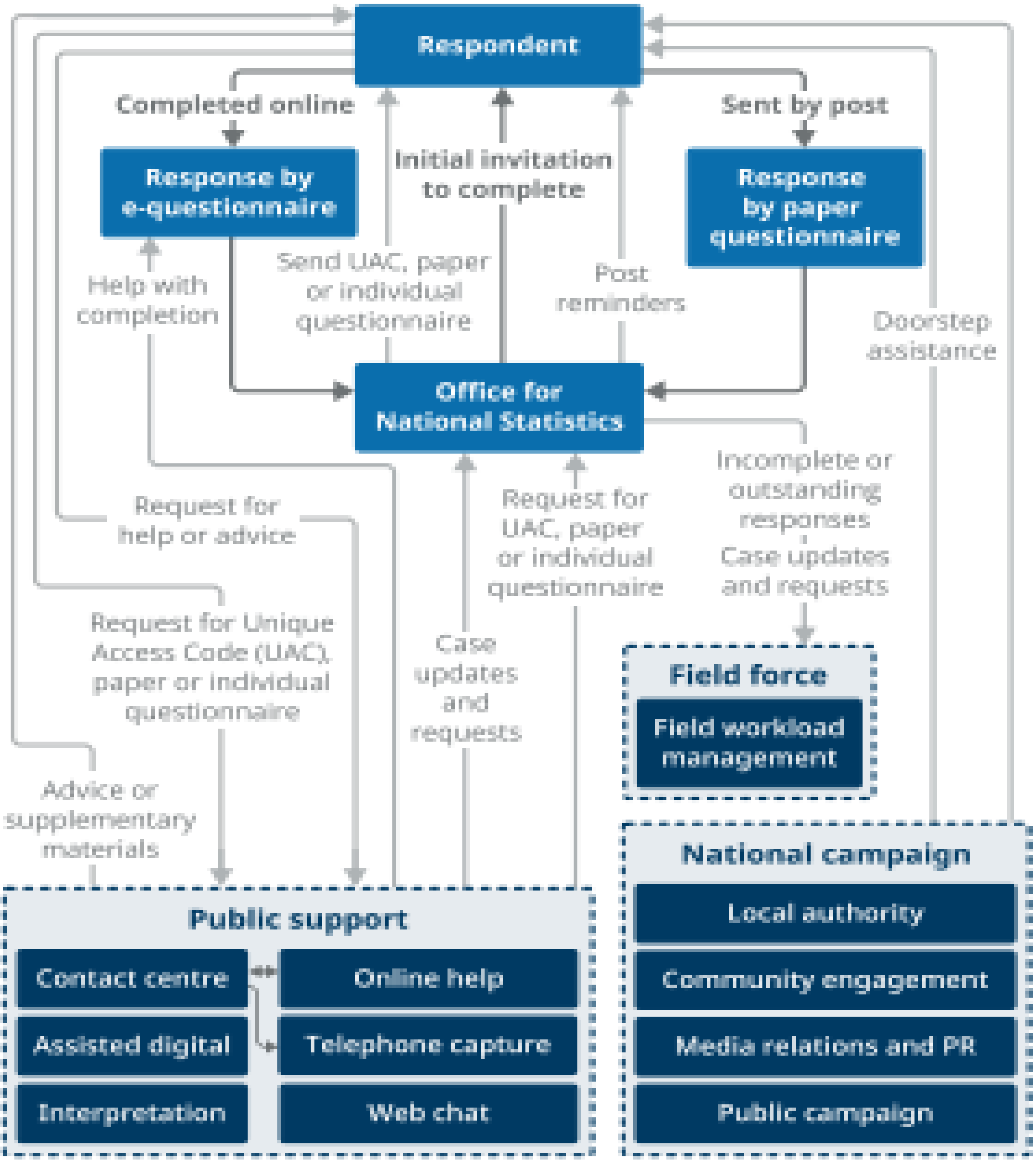
The 2021 UK Census was managed by three agencies: the Office for National Statistics (ONS) for England and Wales, the National Records of Scotland (NRS) for Scotland, and the Northern Ireland Statistics and Research Agency (NISRA) for Northern Ireland. in this report, I focus more on the ONS.
About 97% of all households completed the census. Census data were collected through online submission, Paper Questionnaires, and field visits. 89% of households completed the census, and 20,000 trained field officers were deployed for field visits. Census information was available in 49 languages. The population of England and Wales was 59,597,542 as of 21 March (ONS, 2025).

Data Collection
ONS mailed paper questionnaires and also online access codes sent via post. Electronic questionnaire was completed on the ONS portal and submitted directly to the system, while the paper completed questionnaire forms were returned via Royal Mail which and scanned by accredited suppliers, field enumerators visit for non-responding areas, personal data protected under census Act confidentiality rules (ONS, 2025)
The ONS mitigated bias in different angles, but I am concerned that the online system was well better structured and error-controlled than the paper questionnaire, which had the two imbalances of quality measures. Imputation of incomplete records can misrepresent the real data attribute of some particular records, which can bring bias to the Census. Census data are not actually counted, but rather estimated, which means the census is not 100% accurate. Therefore, if Ai model are trained with 100% census data there is a possibility of underperformance.

Data professionals must hold on to the ACM code of ethics for any action about to be taken and make sure that the purpose of every data gathering must be to improve society, business and security and course no harm to the individual now or in the future.

Appendix

Illustrative Diagram of How data were collected (ONS, 2025, p. 94)


Respondent
Completed online
Sent by post
Response by e-questionnaire
Initial invitation to complete
Response by paper questionnaire
Help with completion
Send UAC, paper or individual questionnaire
Post reminders
Doorstep assistance
Office for National Statistics
Request for help or advice
Request for UAC, paper or individual questionnaire
Incomplete or outstanding responses
Case updates and requests
Request for Unique Access Code (UAC), paper or individual questionnaire
Case updates and requests
Field force
Field workload management
Advice or supplementary materials
National campaign
Local authority
Community engagement
Media relations and PR
Public campaign
Public support
Contact centre
Online help
Assisted digital
Telephone capture
Interpretation
Web chat

REFERENCES

ChatGPT (2025) *Response to a question about classification drift and harmonisation in the UK Census*. Available at: https://chat.openai.com/ (Accessed: October 28, 2025).

Ethnicity Facts and Figures (2021) *List of ethnic groups*. Available at: https://www.ethnicity-facts-figures.service.gov.uk/style-guide/ethnic-groups/ (Accessed: October 28, 2025).

GO-FAIR (no date) *FAIR Principles*.

Office for National Statistics (2021) *Privacy statement – Office for National Statistics*. Available at: https://www.ons.gov.uk/census/censustransformationprogramme/privacystatement (Accessed: November 2, 2025).

Office for National Statistics (2022) *Maximising the quality of Census 2021 population estimates*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/maximisingthequalityofcensus2021populationestimates (Accessed: October 25, 2025).

ONS (2023) *Quality and Methodology Information (QMI) for Census 2021*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/qualityandmethodologyinformationqmiforcensus2021 (Accessed: October 24, 2025).

ONS (2025) *Census 2021: General report for England and Wales*. Available at: https://assets.publishing.service.gov.uk/media/6850323f29fb1002010c4ece/Census_2021_General_report_for_England_and_Wales.pdf (Accessed: October 24, 2025).

UK Data Service (2024) *An introduction to the UK Data Service*. Available at: https://ukdataservice.ac.uk/app/uploads/introtoukds2024-11-07.pdf (Accessed: November 3, 2025).