NILS 2021

ALPHA TEST REPORT

March 2024

1.0 Introduction

The aim of the Northern Ireland Longitudinal Study (NILS) 2021 project was to update the current NILS database with Census 2021 data extracts. This project extends research data coverage to 40 years thus significantly enhancing the continued relevance and value of the NILS data for demographic, health and socio-economic research.

2.0 Background

The NILS contains linked census and life events data for approximately a 28 percent sample of the population of Northern Ireland. The NILS-Core sample was drawn in 2001 by selecting individuals with 1 of 104 birth dates from the BSO Health Card Registrations, who were identified as being 'Live'. This data was linked to Census 2001 and Vital Events (births, deaths) information and has since been updated with the further inclusion of Census 1981, Census 1991, Census 2011, marriages, migration, weather, pollution, property and school information.

For the NILS 2021 project, the Census 2021 data was added to the NILS members and the household information was enhanced to refine the list of householders related to the NILS members, thus bringing the NILS database up to date.

The NILS 2021 database was extensively tested before being released to the research community. The first testing stage, the Alpha Test phase, was carried out by RSU staff. This report documents the quality assurance practices applied by RSU to the NILS 2021 database to ensure the data is accurate and reliable.

3.0 Alpha Test Phase

RSU completed the build of the NILS 2021 database throughout August 2023 as the data extracts required for the build were made available from the data providers (Census Office, Vital Statistics and the Trusted Third Party (TTP)).

As the build process took place RSU worked with the data providers to highlight any issues found with the data. The final census data was provided to RSU on 2 October 2023. On completion of the full build a range of data quality checks were then performed by RSU on the NILS 2021 database to determine whether it was ready for use by researchers. There was limited time to complete this Alpha test phase due to the project plan to have Beta project datasets available to researchers from 9 October 2023.

The objectives of the Alpha Test phase were:

- to test the integrity and accuracy of the new NILS 2021 research database
- to use the evidence gathered through a range of checks and tests to inform the decision on whether to proceed to the Beta Test phase

3.1 Scope of Alpha Testing

RSU aimed to test the entire NILS 2021 database with an enhanced focus on the Census 2021 data. RSU applied a series of quality assurance checks to the data as detailed in Annex A. RSU aimed to look for potential issues surrounding the following:

- Verify record counts match expected numbers (against archived versions of the database)
- Confirm all agreed variables are included
- Test linkages between tables
- Examine comparability of Census 2021 data with Census 2011 data
- Examine consistency between Census 2021 data and Census 2011 Census data
- Check for missing data
- Verify range of values for each variable against Census 2021 data dictionary
- Check for duplicates
- Verify aggregations

3.2 Issues Found

Annex B documents the issues found by RSU whilst carrying out the Alpha testing. Queries were relayed to the data providers and any resolutions are included. As noted, some of the items were resolved prior to the Beta testing phase, those outstanding were investigated further by the data providers with the aim of finding a resolution for the full launch of the NILS 2021 database in early 2024.

4.0 Beta Test Phase

Following the Alpha Test phase, the data was released to researchers for Beta testing. Five beta test projects were developed with a small group of invited experienced users of the NILS. These methodological projects assessed key aspects of the data in short clearly defined projects.

Project specific data extracts were made available to all five Beta research projects on 9 October 2023. The Beta test projects aimed to test the NILS 2021 data for accuracy, comparability and consistency though active analytical use and feedback from the test participants would also inform the decision on whether to release the NILS 2021 database to the wider research community. A separate Beta Test Summary Report documents the Beta Test phase.

5.0 Final Build of NILS 2021 Database

Prior to the formal launch of the NILS 2021 database in April 2024, further work was completed to resolve outstanding issues found in the Alpha and Beta testing phases and to add routine updates and additional vital events data to the NILS 2021 database. On completing the build of this final version of the NILS 2021 database, RSU carried out further testing with particular focus on the items that were unresolved prior to the Best Test phase. Annex B details a summary of the resolution to all items found throughout the testing cycle by RSU.

Annex A: NILS 2021 Alpha Testing Checks Completed

The NILS 2021 database consists of the following tables:

CORENILSDATA	ADDRESS_HISTORY	MIGRATION_EVENTS	XAGES	EVENTS
BIRTHSSTATS	DEATHSSTATS	XDEATH_DETAILS	MARRIAGESSTATS	
PROPERTIES	DATA_RELEASED	SCHOOL		
CENSUSP_1981	CENSUSHH_1981	CENSUSP_1991	CENSUSHH_1991	
CENSUSP_2001	CENSUSHH_2001	CENSUS01_RELATIONSMATRIX		
CENSUSP_2011	CENSUSHH_2011	CENSUS11_RELATIONSMATRIX		
CENSUSP_2021	CENSUSHH_2021	CENSUS21_RELATIONSMATRIX		
POLLUTION	WEATHER	WEATHER_AND_POLLUTION		

The quality assurance checks applied by RSU to the NILS 2021 database included, but were not limited to:

ALL TABLES

- Examine the row counts for each table compared to the previous NILS database.

DATARELEASED

- Examine the DATA_RELEASED table to ensure all latest available data is as described.

CORENILSDATA

- Confirm CORENILSDATA has SOURCE values from 200104 to latest available.
- Count of number of females and males in each download to check for similar trend. Ensure no blank values for GENDER.
- Confirm STATUSHIST values are populated.
- Confirm CURRENT_ADDRESS_SOA values are populated.
- Calculate percentage assigned CURRENT_ADDRESS_SOA to check for consistency of distribution across downloads.
- Confirm CURRENT_ADDRESS_XUPRN values are populated.
- Calculate percentage assigned CURRENT_ADDRESS_XUPRN to check for consistency of distribution across downloads.
- Count of DODYEAR across each year to check for similar trend.
- For each DODYEAR, examine DODMMM total counts to confirm all populated and count across each month is as expected based on previous months.

XAGES

- Confirm that the most recent data is included.
- Confirm average ages at each download.

ADDRESSHISTORY

- Confirm that the most recent data is included.
- Examine count of SOURCE.
- Examine count of CHANGETYPE values.

MIGRATION EVENTS

- Confirm that the most recent data is included.
- Calculate percentage assigned OUTOF_SOA2001 to check for consistency of distribution across downloads.
- Examine count of INTO_SOA2001 values and check have 890 possible values.
- Confirm count of CHANGETYPE values.
- Calculate percentage assigned OUT OF_XUPRN to check for consistency of distribution across downloads.

EVENTS

- Confirm all seven EVENT_TYPE_CODE values are populated.
- Investigate the LINKID to look at the number of NILS members involved in each birth event e.g., baby only, mother and baby etc.
- Investigate the LINKID to look at the number of NILS members involved in each marriage event e.g., bride only, bride and groom etc.

BIRTHSSTATS

- Confirm all GROBID's appear in Events table.
- Examine count of births by REGYR to check all years from 1974 to latest available and total counts for each year look as expected based on previous years.
- Examine count of births by OCCYR to check all years from 1974 to latest available and total counts for each year look as expected based on previous years.
- For each OCCYR, examine OCCMONTH total counts to check all populated and count across each month is as expected based on previous months.
- Group MARSTATOFPARENTS variable to check for expected values.
- Group REGDETAILS variable to check for expected values.
- Confirm all new variables named in table (data will be populated for 2022 onwards).

DEATHSSTATS

- Confirm all GRODID's appear in Events table.
- Examine count of deaths by DTHYR to check all years from 1991 to latest available and total counts for each year look as expected based on previous years.
- For each DTHYR, examine DTHMNTH total counts to check all populated and count across each month is as expected based on previous months.

- Confirm all new variables named in table (data will be populated for 2022 onwards).

MARRIAGESSTATS

- Confirm all GROMID's appear in Events table.
- Examine count of marriages by REGYR to check all years from 2005 to latest available and total counts for each year look as expected based on previous years.
- For each MARYR, examine MARMNTH total counts to check all populated and count across each month is as expected based on previous months.
- Examine counts by Type and Venue.

CENSUS PERSON

- Confirm Census Person IDs of NILS members are all in CoreNILSData table.
- Confirm all Household IDs are also in the Census Household table.
- Confirm all Census Person IDs are distinct.
- Confirm that Census Household IDs can appear more than once.
- Confirm count of NILS Members.
- Confirm all NILS members have a NILSID.
- Confirm all records for SEX variables have binary values.
- Confirm count of MARSTAT groups.
- Confirm consistency of variables and values for Census 2021 data against Census 2011 data.

CENSUS HOUSEHOLD

- Confirm all Household IDs are in Census Person table.
- Confirm variables are populated with expected values.
- Confirm consistency of variables and values for Census 2021 data against Census 2011 data.

CENSUS RELATIONSHIP MATRIX

- Confirm all Household IDs are in the Census Person table.
- Relationship variables – confirm valid values and examine counts of each.
- Confirm consistency of variables and values for Census 2021 data against Census 2011 data.
- Explore the types of relationships listed for each household.

WEATHER/POLLUTION

- Confirm all WIDs are in WEATHER_AND_POLLUTION table
- Confirm that specific variables in Pollution are populated to latest available date.
- Confirm all PIDs are in WEATHER_AND_POLLUTION table.

Annex B: Alpha Test Issues Found and Resolutions

Issue	Description	Resolution
1	MarStatOfParents and RegDetails variables in Birthsstats tables have incorrect values.	Updated Birthsstats files provided from Vital Statistics to RSU 18.08.23 Resolved prior to Beta Test phase.
2	Speak Frequency Irish variable absent from the Census Person 2021 table.	Updated Census Person 2021 table provided from Census Office to RSU 24.08.23 Resolved prior to Beta Test phase.
3	Speak Frequency Ulster-Scots variable absent from the Census Person 2021 table.	Updated Census Person 2021 table provided from Census Office to RSU 24.08.23 Resolved prior to Beta Test phase.
4	The format of the household ID values provided on the 2011_SDZ table did not match the format of those in the Census Household 2011 table and therefore the SDZ2021_code and SDZ2021_name from the 2011_SDZ table could not be linked to the 2011 Census Household table.	Updated 2011_SDZ table provided from Census Office to RSU 20.09.23 Resolved prior to Beta Test phase.
5	Census household IDs in the Census Person 2021, Census Household 2021 and Census Relationship Matrix 2021 tables are in a different format to those included in the Census 2011 tables.	Updated Census Person, Household and Relationship Matrix 2021 tables provided from Census Office to RSU 09.02.24 Resolved prior to final launch of NILS 2021 database.
6	Census Person 2021 table contained 25,531 records with blank Household_ID_H2. This resulted in 5,642 records in the NILS Census Person 2021 table with blank CensusHID2. These have a UR_Residence_Type value of 2 (communal establishments). This resulted in no communal establishments being present in the NILS Census Household 2021 table.	Updated Census Person 2021 table provided from Census Office to RSU 02.10.23 Resolved prior to Beta Test phase.
7	All values in NILS Census Household 2021 table for the following variables are -8: CATERAGE_H2, CATERFOR_H2, CE_MANAGEMENT_TYPE_H2, EMANAGE_H2, NATURE_H2.	Updated Census Person 2021 table provided from Census Office to RSU 02.10.23 Resolved prior to Beta Test phase.
8	Census Person 2021 table: SEX_P2 variable. The Data Dictionary states that value 1 is female and value 2 is male. In Census 2001 and Census 2011 value 1 was male and value 2 was female.	Data as recorded in Census 2021.
9	There is inconsistency in the values used for Census 2021 variables compared to 2011 i.e. categorical variables are coded differently.	Data remains unchanged to match the way the data was recorded in Census 2021.
10	Relationships in the Census 2021 Relationship Matrix table are only included one way unlike in the Census 2001 Relationship Matrix and the Census 2011 Relationship Matrix.	Updated Census Relationship Matrix 2021 table provided from Census Office to RSU 22.02.24 Resolved prior to final launch of NILS 2021 database.

11	It looks like the Relationship_R2 variable in the Census 2021 Relationship Matrix table should be defined as Relation P1 to P2 rather than Relation P2 to P1.	Updated Census Relationship Matrix 2021 table provided from Census Office to RSU 22.02.24 Resolved prior to final launch of NILS 2021 database.
12	There are 215 NILSIDs in the Census Person 2021 table that are not in the Core NILS table.	Resolved prior to final launch of NILS 2021 database.
13	Are the -8 values (defined in the data dictionary as No Code Required) consistent throughout the data? It is unclear in some instances whether this is to account for missing/blanks or if it is a question that didn't require an answer from specific groups. For example, Sexual_Orientation_DVO_Detailed_P2, has -6 uncodeable, -9, -8 No code required. This led RSU to question the absence of 'missing' values in the 2021 data that appear to be present in previous census extracts.	Further clarification added to the data dictionary. Resolved prior to final launch of NILS 2021 database.