

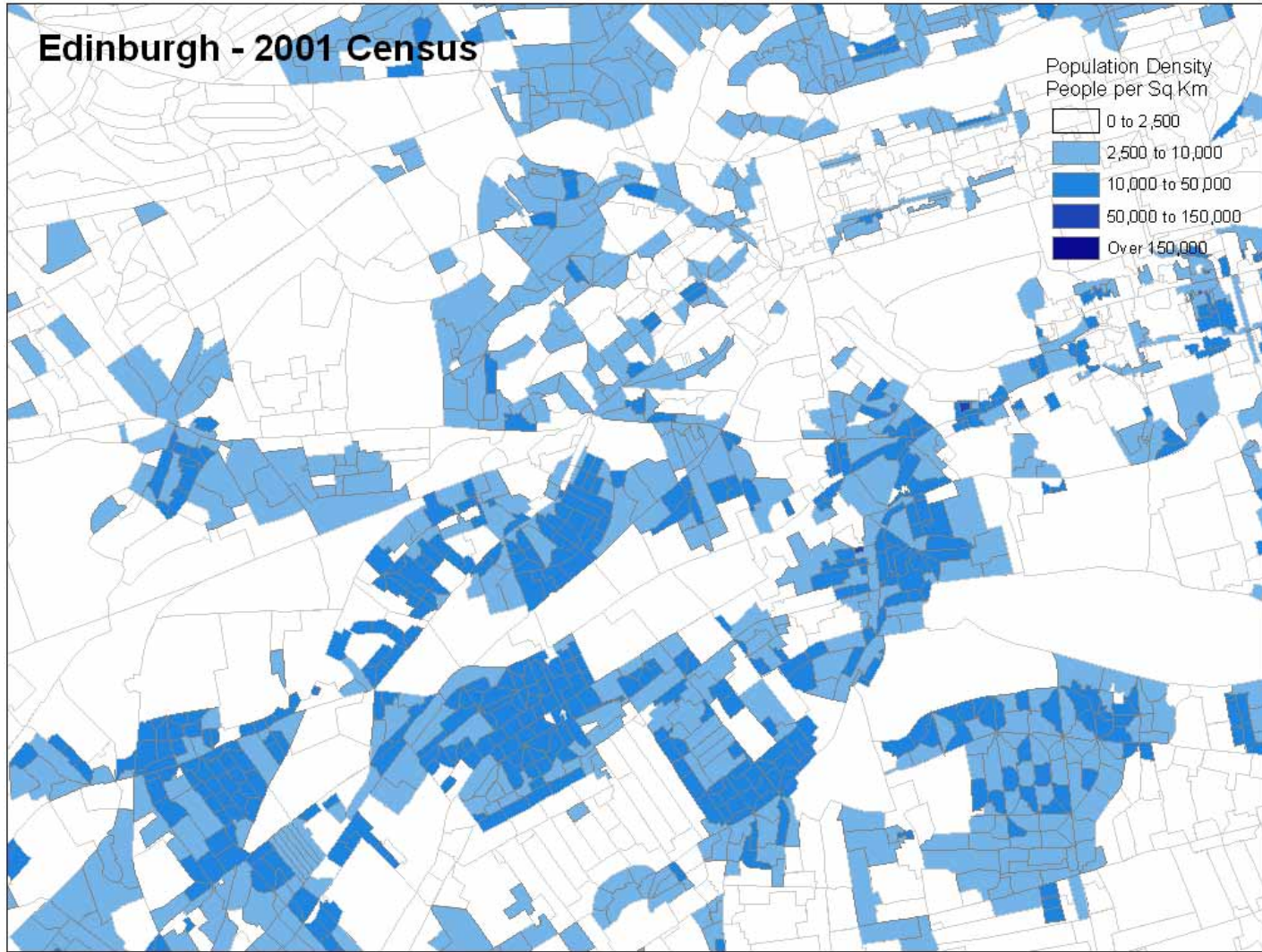
Census Data Quality Assurance

17 May 2010

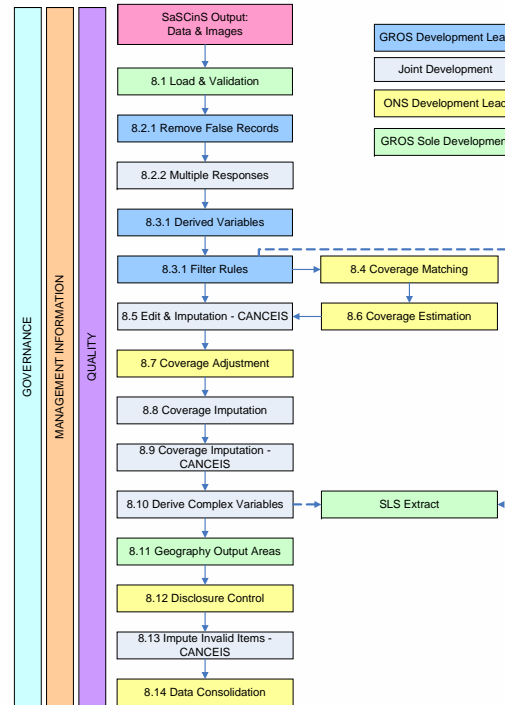
Types of Quality Assurance (QA)

- ▶ Quality assurance of captured and coded data
- ▶ Quality assurance of downstream processes (including data security and integrity)
- ▶ Quality assurance of population counts, variable distributions and other checks on the final data set (e.g. mapping of population densities, quality of workplace addresses)

Edinburgh - 2001 Census



DownStream Processing (DSP)



QA Timetable

Dates	Tasks
August 2009 – April 2011	Testing and improvements on downstream processing system
May 2009 – October 2010	Detailed specification of functionality and checks of the Data Quality Management System (DQMS), including comparator data required
March 2010 – December 2010	Tolerance and diagnostic range methodology devised and built into DQMS
May 2009 – April 2011	Analysis of comparator sources and identification of data quality issues
January 2010 – April 2011	DQMS – IT development and testing
May 2011 – October 2011	Quality assurance of captured and coded data
January 2012 – May 2012	Quality assurance of population counts at local authority level during live running of DownStream Processing (DSP)
August 2012 – December 2012	Detailed demographic quality assurance (on coverage imputation at lower levels of Geography), quality assurance of variable distributions and other checks

QA plans

- ▶ Data Quality Management System (DQMS) with pre-planned analyses to make maximum use of the time available
- ▶ Ability to drill down or carry out ad-hoc investigations as required
- ▶ Use of appropriate comparator data in the DQMS to highlight major differences

Pop. Count
SO1002348

Age group	Rehearsal 09 Count	COMPACTOR	Absolute Difference	% difference	Lower Tolerance %	Lower Bound	Upper Tolerance %	Upper Bound	% diff from lower bound	% diff from upper bound
0-4	24	11	13	118.2	10	9.9	10	12.1	142.4	98.3
5-9	24	42	-18	-42.9	10	37.8	10	46.2	-36.5	-48.1
10-15	24	61	-37	-60.7	10	54.9	10	67.1	-56.3	-64.2
16-19	31	33	-2	-6.1	10	29.7	10	36.3	4.4	-14.6
20-24	25	36	-11	-30.6	10	32.4	10	39.6	-22.8	-36.9
25-29	14	20	-6	-30.0	10	18	10	22	-22.2	-36.4
30-34	18	16	2	12.5	10	14.4	10	17.6	25.0	2.3
35-39	39	50	-11	-22.0	10	45	10	55	-13.3	-29.1
40-44	51	60	-9	-15.0	10	54	10	66	-5.6	-22.7
45-49	45	64	-19	-29.7	10	57.6	10	70.4	-21.9	-36.1
50-54	65	66	-1	-1.5	10	59.4	10	72.6	9.4	-10.5
55-59	90	64	26	40.6	10	57.6	10	70.4	56.3	27.8
60-64	84	83	1	1.2	10	74.7	10	91.3	12.4	-8.0
65-69	112	72	40	55.6	10	64.8	10	79.2	72.8	41.4
70-74	53	53	0	0.0	10	47.7	10	58.3	11.1	-9.1
75-79	43	36	7	19.4	10	32.4	10	39.6	32.7	8.6
80-84	45	38	7	18.4	10	34.2	10	41.8	31.6	7.7
85-89	16	24	-8	-33.3	10	21.6	10	26.4	-25.9	-39.4
90 & over	4	12	-8	-66.7	10	10.8	10	13.2	-63.0	-69.7
Total	807	841	-34	-4.0		841		841	-4.0	-4.0

Current Progress

Use of Rehearsal Data

- ▶ Data has been used to test the DownStream Processing (DSP) stages that have been completed. Improvements have been made to the processes.
- ▶ Early QA stages were tested on rehearsal data (Load and Validation and variable distributions).
- ▶ Rehearsal data is currently being compared to other sources to assess their use in the QA process.

Current Progress

- ▶ Consultations with Analytical Service Divisions within Scottish Government to identify comparator sources and to agree involvement in providing topic knowledge should issues be discovered
- ▶ Close collaboration between General Register Office for Scotland (GROS), Office for National Statistics (ONS) and Northern Ireland Statistics and Research Agency (NISRA) to share knowledge

Ongoing Areas

- ▶ Detailing of checks to be carried out and building of the Data Quality Management System (DQMS)
- ▶ Further analysis of comparator data and preparation of estimates and tolerances to be used in the DQMS
- ▶ Continued testing of the DownStream Processing (DSP) steps when completed
- ▶ Local authority involvement

Local Authority Involvement

Aims

- ▶ To inform about data processing and quality assurance
- ▶ To consider other comparator data sets
- ▶ To gain knowledge of local issues in preparation for quality assurance and for investigation of data anomalies

Questions?