MAKING SENSE OF THE CENSUS

Hundreds of computers will be put to task in handling the massive job of collating and sorting millions of census forms. But how will it work? *Express* lifts the lid on the biggest information technology operation the country has ever seen...

On Sunday 21 April a-once-in-a-decade event took place. The biggest single data processing operation the UK has ever seen, the 1991 census involved the distribution and collection of a staggering 23 million forms.

Other statistics are equally mind-blowing, 135,000 people were involved in the census, including 2,000 temporary staff at the Government's Glasgow centre alone. The forms, which will not be declassified for 100 years, will require 12 miles of racking storage space. Government officials would not say what was the precise cost of census, but agreed that it ran into millions of pounds.

But it is the computer operation which is most fascinating. Over two billion key-strokes and 65,000 megabytes of on-line storage were involved. 80 BTOS workstations, 300 PCs and an Amdahl 5990 mainframe with 192 megabytes of main memory were included among the hardware. The software, which was designed especially for the census, is based on Model 204, an inverted list database designed in the States.

• Staff at the Government's Glasgow office input information from the census forms. Careful planning and raw computing power are being called upon to make the two year collation experience run as smoothly as possible.

It contains 40,000 lines of code written in the Model 204 fourth-generation language and took 18 months to compile. Most of the software will be discarded after the project is completed, in about seven months. It seems like an awful waste, but by the time the next census comes round in 2001, it will have been long obsolete.

All the data processing is being done at the Government's Titchfield, Hampshire office, while most of the data entry will be undertaken in Glasgow. The British Computer Society (BCS) has even been called in as a security consultancy.

TAKE YOUR TIME

Although the first cataloguing of the general population took place as long ago as 1086, when William the Conqueror commissioned the Domesday Book, the first proper censuses didn't begin to evolve until the 17th Century. Before that, inventories were made for specific purposes, usually to find out who was supposed to be paying taxes, doing military service or forced work.

As such these were not censuses; a proper census involves the accumulation of statistics on anonymous individuals, ie, the information gathered by government inspectors is supposed to be used only to chart current trends and predict future ones. The compiling of the Domesday Book caused great popular resentment, and the receptions other inventories received were hardly any better.

COMPUTERS IN THE FRAME

The first time a computer (of sorts) was used in any census was the 1890 US Census when Herman Hollerith used a computerised counting system. Hollerith devised a system of perforated cards and mercury-filled trays to mechanise the laborious counting process. Although this was extremely primitive by today's standards, at the time it was revolutionary and cut the time needed to compile the data by two thirds. Later, Hollerith joined with two companies to form Computing Tabulating Recording Company, which grew into IBM.

Needless to say, the census was as far removed from Hollerith's as was the first US Census from the compilation of the Domesday Book, but the difference is not just a technical one; the Government has made secrecy and security of information its number one priority.

The message to any would-be hackers itching to pit their wits against the system is "forget it". Although computer staff will be able to access information internally, the database is not connected to a public network.

As far back as July 1988 the Government announced a review of the data protection confidentiality and security arrangements for this year. The Data Protection Registrar supported the idea of an independent review, and in its report, the BCS urged that security arrangements be monitored independently, from preparation to processing and storage.

LEGAL PROTECTION

As well as splitting the mainframe up into several virtual machines, all data transmitted over the Government data network between Glasgow and Titchfield will be encrypted.

With the exception of postcodes, names and addresses will not be kept on the computer at all. There are also legal safeguards as well as technical ones. Under the Census Act 1920, it is unlawful to disclose the name or full address of any individual or household; the Government has also given an assurance that access to information will be limited strictly to staff who need it.

Since the last census in 1981 there have been many new developments in computing and in the Government's own set up. These have all been considered by the BCS and include improvements in the technical side of processing which allow massive volumes of data to be kept permanently on the computer, increased (internal) on-line access through the Government's own network, the growth of the "hacker phenomenon" and the implementation of the 1984 Data Protection Act. ■

FUN WITH FIGURES

So, what will the new census tell us about the UK that we don't know already? If you didn't fill out the form personally, ask the head of the household what kind of questions it asked.

Here are a few recent statistics which are likely to be updated, reflecting on current trends:

- In England and Wales the average age of marriage for men is 28, for women, 25.
- In 1985, 13% of all families were single parent families.
- There are 1,053 females to every 1,000 males
- Women out live men by 77.7 years to 71.8 years.
- In 1987, out of 22,272,000 dwellings, 64% were owner occupied.
- We are served by 128,983 doctors, but only [illegible] dentists

HOW INFORMATION TECHNOLOGY HELPS

In the hot summer of 1987 Britain was preparing itself for the Lager Lout. In towns such as Woking, Haverill and Havant young llads, fuelled by liquid dynamite, went on the rampage.

But it wasn't a spontaneous thing. The police knew that a trend would appear, that there would be copycat incidents.

A team of data analysts were called in to study the social set-up of those three towns. They asked the census computer which other towns showed similar characteristics. The computer pumped out names like Guildford and Gravesend.

The police were ready when, within a matter of days, similar incidents broke out in those very towns.

WHY ALL THE FUSS?

With the exception of 1941, there has been a census every decade in Britain since 1801. The census provides information to assist policy and planning, not only by central and local Government, but by health authorities, businesses and numerous other public and private bodies.

Here are a few of the areas in which census information is put to use:

- TRANSPORT STATISTICS: helps plan road building and public transport facilities, and to estimate how many cars will be on the roads in the future.
- EMPLOYMENT: who works where and does what? Information on people's jobs assists both Government and private enterprise in developing job opportunities and training programmes.
- RACE: Britain is now one of the most culturally diverse nations in the world; a break down of different racial groups will tell us who we are and where we come from.
- HEALTH SERVICES: particularly important for local health authorities to enable them to develop services and facilities for the long term sick and the ever-increasing elderly population.
- HOUSING: is another area in which the compiling of statistics is important for local Government, particularly with the current housing crisis.

BEHIND SCHEDULE

It is a standing joke that Government contracts always overrun their allotted budget and timetable; no figures are available for the total cost of the current census, but it must be astronomical. During the course of the BCS's review of security, much of the hardware and software had still not been installed.