10. Computer processing

Computerisation of Census data

One of the main new technical features in forthcoming Censuses will occur in this section: the **use of high performance scanners** [1]**to capture the questionnaire images and, subsequently, to apply OCR techniques;** not only using marks (as in the 1990 Buildings Census), but also handwritten numbers and letters.

In particular, a mixed type method will be used: for the literal manuscripts (in printed text), where the effectiveness of recognition is less, a subsequent comparison will be carried out between the *values recognised* and the real image; for greater efficiency, this comparison will be selective, only checking the literals whose recognised values are not found in specific auxiliary dictionaries (geographic, occupations, activities) or whose recognition reliability index is low. Only where there is a discrepancy between the recognised and real values will it be necessary to enter these values using a keyboard. This will result in a speeded up process in comparison with the traditional method.

The use of scanners will also mean a significant saving to the costs of storing Census documentation by allowing the destruction of papers almost immediately: the image will be much more useful and easier to look after and this process will cover one of the most complex aspects of the protection of Census data.

Computer processing procedures

Computer processing procedures (filtering, coding and tabulation processes) that should be applied to Census data are strongly conditioned by the enormous amount of information to be processed. Thus, procedures, above all those with a certain manual component, which operate perfectly in a sample survey, can be completely inadequate for these Censuses.

On the other hand, users repeatedly request a substantial reduction in the time they have to wait for the different Census results to be published[2]. Both factors coincide in the fact that, apart from guaranteeing an acceptable level of quality, **Census procedures should, above all else, be quick.**

Computer processing in future Censuses will also be conditioned by the need to increase coordination between Autonomous Statistics Institutes and the INE with regards this work, with the aim of achieving a more rational use of resources[3], thus achieving savings on costs and deadlines.

In this way, the new collaboration model consists of a division of functions between the INE and the Autonomous institutes, in such a way that each computer stage is carried out only once. Those Autonomous Communities that are interested in taking part in this aspect of the Census will take on the supervision of work concerning the computerisation of Census documents from their respective territories, in agreement with general regulations that have been previously agreed and in which initial filtering and coding work will be integrated.

---

[1] The best current models easily exceed the barrier of 100 DIN-A4 per minute.

[2] This is an aim that is mentioned again and again when talking about how to focus each relevant aspect of the Census, which clearly demonstrates that it should be one of the key issues in the next Census operation.

[3] This avoids processes carried out in parallel and in particular, the uncoordinated correction of the same mistakes via procedures that in general are very similar, but which logically do not produce the same results. This model, applied in 1991, was a good starting point for collaboration in this area of work, but should be substantially improved for 2001.

The criterion for deciding which processes need to be concentrated on in this area should be taken from the scanned images, in order to provide a better solution. The aim of these processes will therefore be to improve the results of the data capturing procedure (in the sense that it reflects as reliably as possible the answers contained in the physical questionnaires) and should only change their content in exceptional circumstances.

These first processes will be extremely important, because when carried out adequately, they will leave the information partly filtered and coded and of a level of quality that means that subsequent processes will have a minimum manual element.

When this stage is finished, the INE, who is in charge of the data capturing stage, will send a results file to each participating Autonomous institute, including the internal stage of consolidating all register data in all files. As a result of this file, it will be possible to disseminate results previews, always indicating the provisional nature of the figures. Some institutes will subsequently take on intermediate filtering and coding work (the main task being to code the literals that have not been coded in the scanning centre) and, in any case, the INE will ultimately take charge of carrying out the final, aforementioned processes. The minimal manual element to this final filtering work will considerably help in achieving the reduction in deadlines that we are aiming for.

The main aim is to ensure that the most important and well-known product of Censuses, the population figures, are published during the middle of 2002 and that the other Census information is disseminated throughout 2002 and 2003. In this way, by the end if this year, and once the basic Census dissemination programme has been completed, all efforts can be concentrated on making the most of the enormous potential contained in the Census data.