Unintentional duplication of research

A survey revealing instances of belated discovery of information in the literature leads to an estimate of what duplication costs

by John Martyn, Aslib Research Department

In a letter in *New Scientist* (Vol. 19, p. 148) the rate of duplication of research among scientists was said to have been estimated at about 10 per cent. The first systematic attempt to *assess the size of this problem*, carried out by the Aslib Research Department, suggests the actual figure is more than double this, and the consequent cost to the nation to be measurable in millions of pounds.

We put questions to 647 scientists engaged in industrial, academic or government research, including chemists, physicists, biologists, psychologists and mathematicians. One question was: "Have you, during your current research, discovered in the literature information which you wish you had had at the beginning of your project?" To this, 144 of them (22 per cent) replied that they had. Many had made more than one such find, so that the total number of instances was 245.

The scientists indicated the relative importance of these finds by placing each in one of four categories, and their answers are summarised in Table 1. Duplication or waste of work on the scale shown there cannot be ignored.

The first category, of simple duplication, is quite obviously important. Moreover, it seems unlikely that the losses or inconvenience represented by the last three categories were regarded by the scientists concerned as unimportant, since the respondents were usually able to identify their finds by an exact literature reference and in every instance, when asked how the information was discovered, they remembered.

Here we have the useful information that was discovered in the end, but discovered too late. For this sample, it represents an ascertained minimum loss. We can be confident that further useful information which had been missed earlier remained to be discovered, though we cannot measure this "iceberg".

The dimensions of the known loss are sufficient to cause concern and justify an attempt to assign a cash value to it, or at least to determine the order of magnitude of the loss. We cannot estimate the cash loss associated with the need for changes in the plan of research (second and third groups in the table) so we must concentrate our attention on the first and last. There were 88 projects (*not* instances) in these two groups—14 per cent of the total of 647 projects. Some sort of cash loss (scientists' time, in this context, is money) can be inferred to have taken place in at least 14 per cent of all the projects, because of a failure to find published information sooner. However, not all of this information could in fact have been obtained in published form in time to be useful, because it was found that about one-third of it had not actually been published when the project started, although some of it was awaiting publication. Consequently losses due to avoidable late discovery of published information are associated with only about 9 per cent of the projects (two-thirds of 14 per cent).

We cannot translate this figure directly into cash terms, but by making some reasonable assumptions we can arrive at a rough estimate. A modest assumption is that 10 per cent of the funds allocated to the projects mentioned might have been saved had the relevant information been found earlier.

The information found	Number of instances reported
—revealed that their research unintentionally duplicated other work	43
—would, if previously known, have caused them to plan their whole research differently	36
—did, in practice, cause an alteration in the plan of research	60
—would, if previously known, have saved time, money or research work	106
	245

TABLE 1. *Information found "too late" by research scientists.*

It may be a high estimate in some of the cases where "time, money or work could have been saved", but it is almost certainly too low in the cases where duplication, partial or complete, is reported. It seems a conservative guess, therefore, that 0.9 per cent (10 per cent of 9 per cent) of the money spent on these researches could have been saved if all the published information which was ultimately discovered had been discovered in time.

In 1962 the expenditure on research and development in the United Kingdom was about £640 million. Results derived from the sample, if applied to this total, give a figure of about £6 million (0.9 per cent of £640 million) and this is the minimum amount we estimate to have been spent unnecessarily because the published information was not discovered earlier. Assuming that the average total cost of maintaining one research scientist, with his supporting services, is £8000 annually, this loss is equivalent to paying about 750 scientists to do nothing. (It is worth considering whether half this number, if employed as information scientists, or technical librarians, could not have reduced the loss to a much smaller amount).

These estimates are, admittedly, crude, but they are based on ascertained facts. There are grounds for believing that they are too low. Firstly, they contain no allowance for those instances of finding information which changed the course of a project, or would have changed it had it been found possible—the second and third groups in Table 1. These may well have entailed financial loss of some kind—and they apply to 8 per cent of all projects. Secondly, the figures relate only to information which was actually discovered; almost certainly there also exist items of useful information or evidence of duplication which were *not* discovered, entailing further loss. Thirdly, we have assumed that only 10 per cent of the cost is wasted when research is duplicated. It can be, and in some cases is known to have been, nearer to 100 per cent. The true figure for annual avoidable losses could easily be double the estimate of £6 million.