

What to do with extraneous data

This journal intends to follow some others in making available data that do not strictly belong to the scientific papers which it publishes. But that could change.

WHEN is a scientific research report, called a "paper" since the genteel eighteenth century when papers were "read" (out loud, before a small group of interested and disinterested parties), complete? When a fellow scholar, after a few careful readings of the text, is persuaded that he can so fully understand what has been done, and why, that he could march to his own laboratory and repeat the observations he has seen described with a fair chance of replicating their authors' results successfully. Naturally, this does not often happen. The more complete a paper, the less likely are its readers to strive at its replication. But the principle supervenes: to be credible, a paper must contain the essence of its replication. On that, everybody agrees.

The difficulty in that general statement is that the definition of completeness is like the definition of the length of the ideal piece of string: it depends on what the string is needed for. Most papers published in the journals of record (of which this is one) give well-informed readers a good sense of how they might most usefully set about the task of replication; there are densely printed figure legends that detail how the work has been done, and which are therefore a guide to the ways in which it might be repeated. Sadly, of course, the people best placed to understand these coded descriptions of experimental protocols are most of all interested not in replication but in going one step further.

At the same time, it is conventional that the author or authors of a paper will select from the whole corpus of the data at their disposal the few items that will help to prove their case. In palaeontology, for example, people will describe a single specimen in immense detail, and send a dozen others off to a museum unrecorded, on the indisputable principle that a single well-recorded specimen has greater value than sketchy descriptions of a greater body of material.

In human genetics, it is now commonplace to illustrate a general principle by a single example of a familial pedigree — and to put the others in a bottom drawer. In molecular biology, people seem to know that journals will no longer print nucleotide sequences *in extenso*, with the result that they offer for publication only the sub-sequences that, perhaps by comparison with other sub-sequences, will convince percipient readers that there is a case to be made for a particular molecular

mechanism.

Potentially, these practices are a means by which rafts of data may be permanently lost to the research community. Everybody knows that the best that can happen to data that have led nowhere is that they should be "archived", which means in practice that they should be recorded on magnetic tape and sent to a national or international centre for storage indefinitely.

But storage is expensive, and also hazardous; the General Accounting Office in the United States estimated the other day that 10 per cent of the data gathered by the US National Aeronautics and Space Administration (NASA) had been lost because the magnetic tapes concerned had been kept at too high a temperature, or simply misplaced. And that in a field in which the mechanics of data-gathering have been given close attention.

That is why this journal has embarked on a scheme for assisting with the publication of data that would otherwise be buried in people's desk-drawers. For the past few weeks, editors responsible for particular manuscripts submitted for publication have been offering to their authors the opportunity of submitting extra material for publication by a novel route; readers who get to hear, usually by means of an announcement in *Nature*, that there is something extra on offer will be invited to write in, and will be offered a simple photocopy of whatever the author or authors may have supplied, in the first instance without charge.

The immediate beneficiaries of this scheme should be the molecular biologists generating nucleotide sequence data, only some of which is relevant to the conclusions they wish to draw. Why should they not send in the extra data on the understanding that it will be published informally to those who wish to know what it consists of, alongside the more general circulation of their main message?

But there are other fields in which a service such as this should be helpful not merely to the world at large but to those who publish scientific papers — astrophysicists are forever finding that they cannot publish lists of all the stars they have observed, seismologists are at a loss to know what to do with compilations of seismic data, ocean drillers' logs of the constitutions of deep-sea cores hardly ever see the light of day. But why should they not be generally available?

That is the principle on which *Nature*, in

its dealings with those who kindly contribute its scientific articles, will in future function. Those contributing articles accepted for publication will be asked — and are hereby invited — to submit for informal publication such supplementary information as there may also be to hand. *Nature* will undertake to make that supplementary information permanently available. Often, there will be occasions when referees will suggest that the authors of disputable or even contentious articles should, in prudence, put supporting evidence through the same route.

For the time being, at least, none of this should suggest that anything will change. Authors who say that their supplementary information seems to them a perfectly satisfactory basis for a second bite at the same cherry in another journal will not be pilloried on that account. Nor, for the time being, will it be required of authors that they should replace long descriptions of experimental techniques by references to some standard handbook, on the understanding that a full account will be distributed to other parties sufficiently interested to ask for it.

For the time being, the mechanics of the process will be simple, and as described above. Contributors and would-be contributors need not now change their ways. But in the long run, they may discover a novel pressure breathing down their necks; not that expected of even more recalcitrant editors, but that stemming from the inner voice that asks whether this or that typescript will have a better chance if some of the data it reports are tucked into neat supplementary data sheets. If a large proportion of this journal's present contributors followed such a course, *Nature* would be able to publish a greater diversity of material.

The more important objective is that editors and contributors should come to a more direct appreciation of each other's often conflicting interests. Most editors have an interest that most readers should understand as much as possible, most contributors that what they have to say should be unassailable even at the cost that it is unintelligible.

That is why contributors (and readers, often the same people) are asked to give this venture an honest trial. Eventually, of course, supplementary information will be distributed electronically, through an electronic database. But that is light-years away.

John Maddox