

Do Technical Reports Become Published Papers?

By Dwight E. Gray and Staffan Rosenberg

When the study discussed here was conducted, both authors were associated with the Library of Congress in Washington, D. C. Dwight Gray has since joined the National Science Foundation, while Staffan Rosenberg is still with the Library of Congress.

SINCE World War II the technical report has come to occupy a position of considerable importance among media in which results of scientific research first appear. The estimated 50 000 or more documents of this kind which are issued each year—particularly by laboratories engaged in defense-related research—constitute the first appearance in print of much of the newest and most important scientific information being developed today. Almost half of these reports carry no security classification. Yet, for the most part they are not covered by the standard abstracting and indexing services. This is partly because the enormous complexity of the scientific report picture would make systematic coverage very difficult and partly because it would be almost impossible to assure readers of the abstracts of access to the parent reports.¹

The information dissemination problem posed by this situation would largely be solved if substantially all of the significant information presented in these scientific reports were to find its way promptly into conventional scientific and technical journals. In the study described in this paper we investigated a representative sample of unclassified documents of this kind to determine the extent to which journal publication of report results actually had occurred.² Results of the investigation seem to support the following conclusions—evidence for which is presented later in this paper:

1. That the results of this study probably can be generalized with reasonable reliability with, however, any error probably being on the optimistic side.

2. That it makes almost no sense to inquire merely whether or not a given report has been published; one must ask rather as to the publication fate of the significant scientific information in it.

3. That 60–65% of unclassified technical reports in general contain publishable information.

4. That for about half of the technical reports that contain publishable data, all such information is published within 2–3 years.

5. That for about a fifth of the reports that contain publishable data, none of the information is published

at least for several years, and much of it probably never appears in conventional printed form.

6. That a report may contain no publishable information and still be very valuable and well worth preserving, at least for several years, in some readily-retrievable form.

In considering, in the balance of this paper, the data on which the above conclusions are based, we shall first describe the selection of the report sample, then outline the procedures followed in making the study, and finally discuss the results as they pertained both to extent of publication and to time lag between issuance of the report and appearance of the journal paper which included the report data.

Selection of the Report Sample

THE study discussed here was preceded by a pilot survey,³ made in 1952, of 100 reports—25 each in four scientific fields. The experience gained in this preliminary investigation was applied in the later one where the report sample was believed to be sufficiently large to permit generalizing from the results. The gross sample of documents studied comprised the 2500 unclassified technical reports abstracted during the first six months of 1952 in the *Technical Information Pilot* (TIP)—a reports abstracting bulletin published for the Office of Naval Research by the Navy Research Section,⁴ Library of Congress. Our reasons for choosing this particular group of documents were principally:

1. Total TIP listings in any half-year period constituted a large and quite representative sample of reports from a wide variety of report-issuing agencies, including military laboratories and contractor laboratories in universities, research foundations, and industrial organizations.

2. The period January–June 1952 seemed both far enough back for the publication history of the reports to be fairly well established and recent enough for the authors probably to remember having written the documents.

3. Coverage of these reports by TIP meant that a catalog card (usually with an informative abstract) was available for every document, insuring accurate identification to the author of the report about which he was being questioned.

¹ For a description of the National Science Foundation program directed toward improving the general availability of unpublished Government scientific reports, see Gray, Dwight E., *College and Research Libraries*, 18, 1, 23–27, January 1957.

² The study was conducted by the Technical Information Division, Library of Congress, and was supported principally by the National Science Foundation; supplementary personnel support was provided by the Library and by the Armed Services Technical Information Agency.

³ Conducted by Mrs. Helen Brownson, National Science Foundation.

⁴ Predecessor to the ASTIA Reference Center.

Procedures Followed in the Study

PRELIMINARY investigation, amply substantiated by subsequent findings, made quite clear the fundamental fact which is the second of the general conclusions listed above—namely, that *it makes almost no sense to inquire merely whether or not a given report has been published*. Seldom does one report lead to exactly one published paper. In some cases, a single journal article will include the important results from a number of reports; in others, a single report may provide data for several papers; in still others, results from one or more technical reports are integrated with previously unreported data to constitute one paper for publication; and so forth. This situation appreciably complicated our study procedures because it meant that we had to attempt to determine whether a document contained significant scientific information and, if so, the extent to which that information had been published.

Basically, we followed a questionnaire procedure, sending our inquiries to the authors of the reports for whose contents we were trying to establish the publication history. However, as a result of our initial study of the catalog cards for the 2500-report sample, we identified four groups of documents that we decided to omit from the questionnaire survey; this decision was made in consultation with an advisory panel from the agencies supporting the study. In the first such group were 64 reports whose contents we already knew had appeared quite fully in technical journals. The second group, containing 335 items, qualified for investigation in every respect except that obtaining reliable information about their publication histories would have been very difficult, if not impossible; for some of these no personal authors were given, complete identifying information was missing for others, and a number were themselves compilations of separate reports. We saw no reason to believe that the average publication picture for these documents would be appreciably different from that of the more than a thousand reports we could investigate fairly easily; consequently, these were omitted from the questionnaire phase of the study.

In our third excluded category were 154 theses, foreign documents, translations, bibliographies, and the like. Although there had been good reason for announcing these in the TIP bulletin, we felt they were not US technical reports as that term commonly is used and that their inclusion in the questionnaire survey might somewhat cloud the results. Finally, there were 295 reports whose contents, although technical, seemed quite definitely to fall outside the scope of material normally accepted by editors of scientific journals; these included chiefly detailed test and evaluation reports, extensive and detailed compilations of tabular data, and documents on experiments that failed. For these, we arbitrarily made the decision that they did not contain information suitable for publication. (That this conclusion by no means implies the reports were valueless is discussed later.)

Elimination of the above groups of reports left 1652

documents whose authors were queried regarding their publication patterns. For each of these, the writer was sent a questionnaire which asked him first whether or not, in his opinion, that particular document contained "publishable" information; this was defined for him as "information which you think would warrant publication in a scientific or technical journal, book or other generally-available source". The questionnaire then asked those who replied to the first question in the affirmative to answer three other queries, as follows:

1. What percentage of such publishable information has been published, and where? (Complete citations were requested.)
2. What percentage, although not yet published, is definitely scheduled for publication, and where?
3. Why have no steps been taken toward publication of whatever fraction of the publishable information is not included under No. 1 or No. 2?

A catalog card for the particular report was attached to the questionnaire to insure accurate identification of the documents by the author. One thousand and eighteen questionnaires were returned relatively fully answered.

In any given case, one might quarrel with the author's subjective judgment as to whether his report contained information worth publishing. To a considerable extent, however, the same can be said of manuscripts which engineers and scientists submit to journals. We believed it important to use in this study a criterion of "publishability" that would be reasonably comparable with that which applies in the normal technical publishing situation. Author opinion seemed to us to be the closest approximation to this we could obtain without convening a sizable panel of journal editors and review board members from a variety of scientific fields—a procedure that was not practical in this study.

Table 1 summarizes the various subgroups of the gross report sample as they were described above. The first two of these—totalling 1082 documents—constitute what we shall call the "net sample". These are the reports for which substantially complete publication histories were obtained and on which the results and conclusions presented in the following section largely were based.

Table 1. Breakdown of Gross Sample of 2500 Reports

64 reports—contents known, when the study began, to have been quite fully published
1018 reports—questionnaires sent out and returned completely answered
634 reports—questionnaires sent out but either not returned or returned inadequately answered
335 reports—omitted from questionnaire survey only because of obvious difficulties in obtaining data
154 reports—theses, foreign documents, translations, and other publications judged not to come within the scope of "technical reports"
295 reports—contents technical but, in the opinion of the study staff, clearly not publishable
2500 reports—total gross sample

Analysis of Results on Extent of Publication of Report Data

TABLES 2, 3, and 4 represent summaries of three aspects of the questionnaire results—reports that contained publishable information, fraction of such information published, and reasons the remainder was not published.

Table 2. Reports in Net Sample Containing Publishable Information

(As judged by report authors)

	<i>No. Reports</i>	<i>%</i>
Contained publishable information	822	76
Contained no publishable information	244	22
Author unable to decide	16	2

Table 3. Fraction of Significant Data Published from the 822 Reports Containing Such Information

(After 2-3 years)

<i>This fraction of significant information was published</i>	<i>From this many of 822 reports</i>	<i>From this percentage of 822 reports</i>
All	393	48
More than $\frac{3}{4}$	512	62
More than $\frac{1}{2}$	572	70
More than $\frac{1}{4}$	596	73
None	189	23

Table 4. Reasons for No Publication of Information in the 189 Reports

(As given by report authors)

20 reports—although once publishable, data have been superseded
27 reports—research still in progress
64 reports—research completed but manuscript not yet prepared
17 reports—manuscript completed but not yet accepted for publication
14 reports—author simply said he had no plans to publish information
47 reports—miscellaneous other including (a) author did not have time to conform to editorial policies of journal, (b) author believed information sufficiently available in report, (c) author was afraid material bordered on being classified, (d) author could find no suitable journal in which to publish, and (e) author had transferred to a new field and lost interest in old one (this group included one draftee)

In analyzing the data presented above, we tried to keep firmly in mind, first, that unless terms are carefully defined and assumptions clearly recognized any generalization is likely to be nonsense and, second, that even when such precautions are taken, generalizations sometimes can be quite misleading. The several conclusions presented at the beginning of this paper seem to us to be warranted by the data. We shall now repeat

these, accompanying them this time with our reasons for believing them sound.

1. That the results of this study probably can be generalized with reasonable reliability with, however, any error probably being on the optimistic side.

Although the wide variety of originating sources of TIP-listed reports certainly tended to make that journal's coverage representative of technical reports in general, the nature of the bulletin's sponsorship may have resulted in a slight "loading" of the total listings in favor of Office of Naval Research reports. Since, at that time, ONR probably placed greater emphasis on journal publication of report data than did most other contract-administering offices, the extent-of-publication figures shown here may be a little higher than the true averages for all technical reports. With this one reservation, we assumed the findings of the study to be reliable and capable of generalization in arriving at the rest of our conclusions.

2. That it makes almost no sense to inquire merely whether or not a given report has been published; one must ask rather as to the publication fate of the significant scientific information in reports.

Reasons for this conclusion already have been discussed in some detail.

3. That 60-65% of unclassified technical reports in general contain publishable information.

The study showed that of the reports whose publication patterns we were able to establish (i.e., of the net sample), about 76% contained publishable information (Table 2). It seems reasonable to expect that this same percentage would hold for reports in two other subgroups—the 634 for which questionnaire answers were not returned and the 335 whose authors could not be queried readily. Applying the 76% figure to these subgroups as well as to the net sample, we obtain a total of 1560 reports, or 62.5% of the gross sample of 2500 reports, that probably contained publishable information. It is the gross sample whose composition is presumed to be comparable to that of the estimated annual grand total of upwards of 50 000 technical reports.

4. That for about half of the technical reports that contain publishable data, all such information is published within 2-3 years.

The actual figure for the reports in our net sample from which *all* publishable information had appeared in print, as shown in Table 3, was 48%. As the table also indicates, smaller fractions of such data had been published from considerably higher percentages of the documents. The time span of 2-3 years represents a rough average of the period from issuance of the reports covered by this study to receipt of the last completed questionnaire.

5. That for about a fifth of the reports that contain publishable data, none of the information is published, at least for several years, and much of it probably never appears in conventional printed form.

Table 3 shows that for 23% of the reports, none of the publishable information was published within 2-3 years. This figure reduces to about 21% if we eliminate the 20 documents whose authors said the contents had been superseded by later information (Table 4). For the rest, it probably is safe to assume that most data unpublished after 2-3 years will remain unpublished.

6. That a report may contain no publishable information and still be very valuable and well worth preserving, at least for several years, in some readily-retrievable form.

The last subgroup in Table 1—comprising 295 reports—was described as not containing publishable technical information within the framework of this study. These categories of document, however, still may be extremely valuable to scientists and engineers—some for limited periods and others for a considerable length of time. Test and evaluation reports, for example, may have immense short-range use even though they are unattractive to journal editors. The same is true of documents which contain only negative results; such information can be invaluable in preventing costly repetition of experiments that failed the first time. Also, there frequently is need for detailed tabulations of data which appear in reports but are much too extensive for inclusion in a journal article.

Analysis of Results on Time Lag in Publication of Report Data

A NUMBER of studies⁵ have pointed up the fact that substantial amounts of scientific data suffer relatively rapid obsolescence. Thus, the time lag between issuance of a report and the appearance of that document's significant contents in the open literature is a matter of some importance. In asking questionnaire respondents to include complete citations for all publications stemming from their technical reports, we hoped to obtain data from which reliable maximum, minimum, and average time lag figures could be established for our report sample. Here again, unfortunately, the lack of a one-to-one correlation between "report" and "published paper" reared its complicating head. Exhaustive analysis of the 633 questionnaire replies that gave any information of this kind seemed to prove that almost the only conclusion one could draw from the data was that no clear-cut, tidy, once-and-for-all conclusions were possible. Among the facts that can be offered to support this discouraging statement are the following:

1. Reports falling within the sample frequently were single items in series, most of which were issued before and after the period covered by the sample. Data from the entire series then were later included in one or more published papers.
2. More or less the same report data sometimes were published in slightly different form in several journals over an extended period of time.

3. Some reports consisted of compilations of raw data on which an entire series of papers, published at different times, was based.
4. Some published papers concerned only auxiliary aspects—instrumentation, for example—of the reports on which they were based with the principal scientific results having entirely different publication histories.
5. Reports frequently are issued considerably later than the dates they carry; sometimes they are issued somewhat earlier.

These are but five of a number of complicating facts which, in combination, made it virtually impossible to isolate scientifically comparable segments of report findings and determine exactly when each first saw the light of conventional publication day. In other words, a rigorous, statistical, and meaningful accounting of the time lag between issuance of a report and journal publication of its significant contents was not feasible—at least from data obtained in this study. At the most, it seemed to us, our careful study of the questionnaire returns would permit us to state only the following qualitative impressions:

1. Of the significant report material *destined for publication*, roughly half is published within one year of issuance of the report, and
2. Most of the rest of the information *destined for publication* probably finds its way into print within the next two years with a few straggling bits coming along after still another year or two.

The italics in these two statements emphasize that they refer to report data exclusive of the substantial amount discussed earlier in this report which never is published.

On the general question of what happens to the scientific information in technical reports one encounters a broad spectrum of personal opinion ranging from the infra-optimistic view that "everything worth while soon gets published and if it doesn't it isn't worth publishing" to the ultra-pessimistic belief that "enormous quantities of extremely important scientific information are being interred in the tomb of technical reports, never to be resurrected in any usable form". The results of this study—perhaps not too surprisingly—indicate that these are typical members of the great "extreme" family in that neither is realistic, with the facts lying somewhere in between. Certainly, a substantial amount of the scientific data in technical reports does eventually make its way into the open literature—albeit, frequently at all too slow and halting a pace. Just as surely, an appreciable segment never achieves this happy state—or does so after so great a lag that publication is of little benefit to anyone but the author of the paper. On the whole there would seem to be clear justification for the dual expenditure of considerable effort—on the one hand, to encourage authors of reports to publish significant material more promptly and, on the other, to insure the availability of scientific report literature as such for at least several years after issuance.

⁵ Several such studies are summarized and referenced in Herner, Saul, "Technical Information—Too Much or Too Little", *Scientific Monthly*, 83, 2, 82-86, August 1956.