

HENRY VOOS

The Information Explosion; or, Redundancy Reduces the Charge!

The information explosion has been blamed for the backlogs in libraries and information centers and for the inability of these centers and libraries to provide relevant information quickly. It has been used to justify the use of automation. An examination of the literature since the early fifties has shown that the library world has been aware of the problem, but has done very little about it. Some preliminary experiments reveal extensive dual publication between technical reports and journal articles; dual indexing and announcement have also contributed to what is more properly called a paper explosion. Methods of decreasing dual publication are suggested.

PUBLISHING'S WALL STREET gold rush of a few years back was set off by an attempt to capitalize on the "information explosion" and a belief that expenditures on educational materials would continue to climb ever onward and upward. Now with cut-backs in Health, Education and Welfare funds signaling that the education boom is leveling off, and the information explosion sputtering on a slow fuse, it appears that the money men are tucking in their tails a bit.¹

The literature, or the information explosion (these are not identical), has been blamed for the current problems facing librarians and information scientists in handling and utilizing the information that is needed by scientists and engineers. Burton Adkinson and Charles Stearns quote a Systems Development Corporation report which states:

Dr. Voos is Associate Professor at Rutgers University, Graduate School of Library Service. This paper was partially funded by the Research Council of Rutgers University.

Modern information technology has made it possible to place much of the accumulated knowledge of the human race within reach of a man's fingertips, so to speak. But the capacity of executives, scientists, and scholars to absorb information has not increased. Therefore, as the amount of available information grows, there is a parallel need for a more precise capability to retrieve specific data in any area of interest.²

The academic world is now confronted with having not too little but too much information available. The student's problem used to be that he could not find enough material on his topic, whether it was the crisis in the Congo or butterflies in Brazil. Now—with the aid of up-to-date bibliographical tools, abstract services, and indexes—he is surfeited with sources.

Librarians and archivists are also feeling the effects of the publication explosion. They are overwhelmed with book catalogs and reprint edition catalogs, to say nothing of the ever growing size of *Publisher's Weekly*.³

Adkinson and Stearns further elaborate

TABLE 1
CONTENT ANALYSIS: NO. OF MENTIONS

Journal	1964	1965	1966	1967	1968	1969	Average
<i>American Documentation</i>	10.00	15.38	15.78	3.84	16.36	20.00	13.56
<i>Special Libraries</i>	7.14	24.32	13.04	5.55	22.22	20.69	15.49
<i>College & Res. Lib.</i>	9.43	6.89	4.44	4.17	7.01	9.61	6.93
<i>Lib. Res. & Tech. Serv.</i>	8.16	13.79	6.82	4.55	8.57	6.06	7.99
Average no. of times used	8.34	15.09	10.02	4.53	13.54	14.09	

that the problem in library utilization of computers "arises out of a long history of financial starvation of library management. It has been intensified in recent years by the exponentially increasing amount of information that has to be obtained and filed, and the increasing library manpower and space that this entails."⁴ This problem has been used in numerous articles as a launching device or excuse for whatever the author was writing about. To determine whether the use of "information explosion" or words to that effect occurred with regularity in the library literature and that of information science, a content analysis was performed on four American library journals for the years 1964 through 1969: *American Documentation*, *Special Libraries*, *Library Resources & Technical Services*, and *College & Research Libraries*. (See Table 1.) Excluded from the analysis were articles which were bibliographies, reports of committees, descriptions of libraries, and state-of-the-art on such items as reproduction methods.

Although it is difficult to assess the significance of this mention (it is significant at the .05 level using the Friedman two-way analysis of variance by ranks), it is obvious that many professionals in the field are using the "information explosion" in their articles for a variety of reasons.

Librarians today are concerned about something that has been called a paper explosion. This is not the best description of a phenomenon that has been going on for thirty or forty years—perhaps "creeping

neoplasm" would be better. Since it has been going on for so long, it must have been noticed, and something must have been done about it. In the mid-thirties, some people were so alarmed about the paper explosion that they proposed a two-year moratorium on all research and development so they could "catch up."⁵

There are many references concerning the extent of the explosion, some of which are in the bibliography at the end of this article. However, not everyone has agreed that the explosion is as large as has been stated.

During the past 10 or 15 years innumerable papers were written which emphasized in an introductory sentence the important and far-reaching consequences of the so-called information explosion. The use of this phrase seems to imply that the tremendous increase in the number of publications automatically produces a corresponding growth in "information" or knowledge. But is this true?⁶

It is possible that the so-called information explosion is not as serious as has been thought.⁷

. . . despite frequent references in the press and elsewhere concerning an exponential growth in technical literature, there does not seem to be any evidence to this effect.⁸

Most of the library and information science literature tends to concentrate on handling the flood rather than damming (damning) it. One factor that provides evidence that the explosion is really not a flood of *new* information is the extent to which much of what is written is redundant. This paper will try to analyze the redundancy. The problems

seemingly created by the explosion might thus be more deeply analyzed and more fruitfully resolved by studying the input rather than bemoaning the output.

The factor of redundancy can be looked at from at least two sides: the first is the multiple indexing of the same material in different indexes and in different ways; and the second is the republication of the same or similar material in a variety of media.

The more time a literature searcher spends culling through indexes for material that might be relevant, but which has already been scanned by him without his being aware of it, the more time is wasted, and the more it costs to search the literature. J. J. Ebersole in looking at this problem states:

As more indexing and abstracting services are created, as the volume of knowledge increases, and as it becomes increasingly interdisciplinary in character there will be an increasing problem of overlap among these services. The major manifestations of this overlap consist of two or more organizations indexing and abstracting the same document.

He then goes on to analyze some of the journals being indexed and abstracted and states that "an analysis of 17,000 journals covered by 11 of the 18 profession-oriented services in 1961 showed a 50 percent overlap in journal coverage among these 11 services alone." He continues:

Although these costs are dramatic, they are probably exceeded both absolutely and relatively by the duplication existing in the coverage of technical reports. This duplication cost extends not only through the community of indexing and abstracting services, but through the vast complex of libraries and information centers operating in hundreds of companies and government agencies.⁹

This view is also echoed by Alfred A. Beltran:

The open literature (articles and technical papers) listed in *TAB* is also listed in standard commercial indexes available in all technical information centers. Inclusion of this material in *TAB* seems to serve no useful purpose while adding to the duplication already prevalent in commercial indexes. This merely increases the time required by a literature scientist to complete a search as he will encounter the same item several times.¹⁰

A prime attack was made on this by Martyn when he showed that, in certain fields, only an additional percentage point of information is gained by scanning more than the prime abstracting journal. A reinforcement of Beltran's point is that the *Technical Abstract Bulletin (TAB)* and U.S. Government Research and Development Reports (*USGRDR*) both index and cite items by corporate authors, while most commercial indexes use either author or title, but not corporate authors.

Although Beltran claims that the open literature was only announced in *TAB* beginning with Bulletin no. 65-7 (1 April 1965) and although he found an average of 6.47 percent of "open literature unavailable" reports, the author presents his findings based on a random sample from the *TAB* and *USGRDR* for 1961 and 1963-68.¹¹ (See Table 2.)

TABLE 2
REPRINTS ANNOUNCED AS REPORTS

Year	No. in Sample	No. of Reprints	%
1961	1222	152	12
1963	1118	79	7
1964	2852	512	18
1965	630	54	9
1967	71	21	30
1968	53	15	28

It is obvious from the ranges that a more definitive study with larger sample size is required, but the extent of this redundant indexing is certainly evident.

Many of the announced research reports are later published in some form

as articles. This adds the second dimension of redundancy to the explosion. Reprinting, or dual publication, has at times been advocated in order to reach multiple audiences. F. W. Hunt explores the reasons for dual publication: publish or perish, immediacy vs permanence, more depth in indexing of the report vs the journal, assertion vs authenticity, unavailability of space in journals, and the philosophy that a report is not a publication.¹²

To determine whether the redundancy factor could be verified in another way, a random sample (1 percent) of citations was taken from *Chemical Abstracts* for the years 1955 and 1960. Excluded from the sample were patents and non-American articles. The Defense Documentation Center was then requested to run a search on the authors of the articles selected from *Chemical Abstracts*. The search showed that in 1955, 11 percent of the articles were available in report form, and in 1960, 27 percent of the articles were available in report form. It is also interesting to note that in 1955 about 11 percent were theses or extracts from dissertations, and in 1960 only 6 percent fell within this category.

Another form of redundancy is the republication in book form of articles previously available as either journal articles or technical reports. Examples of such republication are the many volumes of "Readings" in various fields. A cursory look at the 1970 *Paperbound Books in Print* (v.14, no. 10) shows that about .3 percent of the 80,000 titles announced is in this category, and about .2 percent of the 245,000 titles announced as hard-bound in the 1967 *Books in Print* fall within this category.

A further example of redundant publishing as a contributory factor to the literature explosion is the republication of the writings of a single author. In the field of library and information science,

books by Taube and Shera serve as examples.¹³ The justification for republishing, as mentioned before, is the extension of the audience. This justification has been used, as well as some of the others previously mentioned, for publishing abbreviated versions of dissertations in journals.

Redundancy as a factor has been brought to the attention of the information and publication fields before. The National Science Foundation reported in 1952 that in a sample of 84 reports in various fields, 56 percent had already appeared in whole or in part, 13 percent were in the process of publication, and 2 percent would be published. In addition, 6 percent were easily obtainable, and 21 percent were unsuitable for publication. Of the 47 reports which had been published in whole or in part, 71 publications have resulted.¹⁴ Note that a factor of 1.5 seems to exist in terms of multiple publication.

In 1956, as a project of the National Science Foundation, the Library of Congress did a survey using a sample from the *Technical Information Pilot* of 1952, to determine the extent of republication or dual publication of technical reports. This report showed that 48 percent of the information in the technical reports containing publishable information had been published in the open literature in its entirety (this is only 21 percent of all the reports published), another 14 percent of the reports had published more than 75 percent of their information in the open literature, an additional 8 percent had published more than half in open literature, and 3 percent had published more than 25 percent in the open literature. This left only 23 percent of unclassified, unlimited information that had not also reached the open literature. The reasons for the non-publication of this 23 percent ranged from the fact that the material in the report had become obsolete, through rea-

sons such as "research still in progress," "research completed, manuscript not yet completed," or "manuscript completed, but not yet ready for publication." Only 1.7 percent of the authors stated that

they had no intentions of publishing. The study warns that the figures might be skewed because the *Technical Information Pilot* operated under an Office of Naval Research contract, that the ONR

TABLE 3

ESTIMATED PERCENTAGE OF SUBSTANCE OR CONTENT OF TECHNICAL REPORTS APPEARING IN SCIENTIFIC AND TECHNICAL PERIODICALS AND TIME REQUIRED FOR PUBLICATION

	Total Percent	Within a Year of Completion of Research	1 Year After Completion of Research	2 Years After Completion of Research
Department of Agriculture	48	15	18	15
Commerce:				
National Bureau of Standards	80	80
Coast and Geodetic Survey	80	70	10
Patent Office	50	25	25
Weather Bureau	60	30	25	5
Public Roads	90	60	25	5
Department of Defense:				
Office of the Secretary of Defense	15	4	7	4
Department of the Army	40	33	5	2
Department of the Air Force	30	10	10	10
Department of the Navy	N.A.	N.A.	N.A.	N.A.
Department of Health, Education, and Welfare:				
Food and Drug Administration	100	75	20	5
Office of Education	11	10	1
St. Elizabeth's Hospital	75	75
National Center for Health Statistics	10	5	5
National Institutes of Health	98	40	52	6
Bureau of Medical Services	75	75
Bureau of State Services (Environmental Health)	86	40	40	6
Division of Accident Prevention	90	80	10
Communicable Disease Center	75	75
Division of Dental Public Health and Resources	80	50	20	10
Division of Hospital and Medical Facilities	95	40	40	15
Division of Nursing	90	90
Social Security Administration	95	75	20
Division of Vocational Rehabilitation	55	46	6	3
Department of Interior	50-90	10-75	25-90	50-90
Department of Labor	80-90	80-90
Department of State: Agency for International Development	80	80
Department of the Treasury: Coast Guard	100	100
Post Office Department	(2)	(3)
Arms Control and Disarmament Agency	(2)
Federal Aviation Agency	5	5
Federal Communications Commission	100	25	25	50
Federal Housing Administration	90	50	25	15
National Aeronautics and Space Administration	25	6	12	7
National Science Foundation	95	65	20	10
Office of Emergency Planning	10	10
Smithsonian Institution	100	10	80	10
Tennessee Valley Authority	50	28	12	10
Veterans' Administration	45	41	4

¹ Under 10.

² Less than 5.

N.A.—Not available.

placed great emphasis on journal publication, and that many of the reports did not contain anything publishable.¹⁵

In 1955, Alan T. Waterman stated that in a sample of 100 reports, 25 in each of four fields, the authors were asked whether the information in the reports had been or would be published in the open literature. 56 percent stated that they had also been published in whole or in part in the open literature, 16 percent were in the process of publication, and 6 percent had been announced in journals.¹⁶

The Elliot report provides some data on time lag and extent of publication of the technical report and the same item published in a periodical.¹⁷ Table 3 is copied from the report.

A variety of recommendations have been made to reduce redundancy in the literature.

1. To reduce the time that a literature searcher must spend scanning the indexes and abstracts, the Clearinghouse for Federal Scientific and Technical Information should cease to announce reprints, even if they are the products of government and government-contracted research and development. Since the journal articles are generally abstracted in the standard abstracting and indexing journals, much time and money could be saved in announcing, acquiring, cataloging, and searching if they were not announced in *USGRDR*.

2. Government contract requirements should be modified so that fewer reports need be issued.

3. The system of refereeing should be tightened.

4. Preprints should be a standard distribution mechanism for many items that later find their way into print (which should not). Moore's survey revealed that 40 percent of the members distributed preprints to as many as 200 colleagues on an average of nine months to one year before publication.¹⁸

5. Make strenuous efforts to eliminate the "publish or perish" syndrome in universities.

The very bulk of scientific publications is itself delusive. It is of very unequal value; a large proportion of it, possibly as much as three-quarters, does not deserve to be published at all, and is only published for economic considerations which have nothing to do with the real interests of science. The position of every scientific worker has been made to depend far too much on the bulk rather than the quality of his scientific publications. Publication is often premature and dictated by the need of establishing priorities. . . .¹⁹

There is sufficient evidence, even with limited sampling, to show that redundancy contributes heavily to the literature explosion. It also provides additional evidence that the explosion is not one of information, but rather one of paper. Additional research, including cost benefit studies on nonpublication might help to solve part of the problem. Rather than finding ways of coping with the explosion, more effort should be devoted to investigating its real causes and its extent in other than gross figures.

REFERENCES

1. Richard Lingeman, "American Notebook," *New York Times Book Review* (Mar. 1, 1970), p.24.
2. B. A. Adkinson and C. M. Stearns, "Libraries and Machines—a Review," *American Documentation* 18:121 (1967).
3. K. M. Glazier, "Paper Pollution," *Wilson Library Bulletin* 44:856 (1970).
4. Adkinson and Stearns, "Libraries and Machines," p.123.
5. M. E. Freeman, "The Science Information Exchange As a Source of Information," *Special Libraries* 59:86 (1968).
6. Gertrude London, "The Publication Inflation," *American Documentation* 19:137 (1968).
7. L. H. Mantell, "On Laws of Special

Abilities and the Production of Scientific Literature," *American Documentation* 17:8 (1966).

8. *Ibid.*, p.13-14.
9. J. L. Ebersole, "An Operating Model of a National Information System," *American Documentation* 17:33 (1966).
10. A. A. Beltran, "Evaluation of DDC TAB from Various Points of View," in Regional Workshop on the Report Literature (Albuquerque, N.M., 1965), *Proceedings . . . Western Periodicals*, 1965, p.84.
11. *Ibid.*, p.83.
12. East Coast Navy Interlaboratory Committee on Editing and Publishing, *Some Views on Dual Publication*, by F. H. Hunt, Oct. 1, 1968. (Monograph no. 5, AD676523), p.2-3.
13. M. Taube, *Studies in Coordinate Indexing*, v.1-6 (Washington, D.C.: Documentation Inc., 1953-1965).
J. Shera, *Documentation & the Organization of Knowledge* (Archon, 1966).
———, *Libraries & the Organization of Knowledge* (Archon, 1965).
14. National Science Foundation, *Preliminary Study of Unclassified Government Technical Reports* (July 31, 1952), 6p.
15. Library of Congress, *Final Report on Study of Publications Stemming from Defense-Related Technical Reports* by D. E. Grey and S. Rosenborg (Dec. 1956), 8p.
16. Alan T. Waterman, "The Future of Report Literature," in B. M. Fry and J. J. Kortendick, *The Production and Use of Technical Reports* (Washington, D.C.: Catholic University, 1955), p.6-7.
17. U.S. Congress, House, 88th Congress, 2d Session, Select Committee on Government Research, *Documentation and Dissemination of Research and Development Results*, Study no. IV (Washington, D.C.: Govt. Print. Off., 1964), p.37.
18. C. A. Moore, "Preprints: An Old Information Device with New Outlooks," *Journal of Chemical Documentation* 5:126-8 (1965).
19. J. D. Bernal, *The Social Function of*

Science (London: Routledge, 1939), p.118.

BIBLIOGRAPHY

The bibliography which follows is not exhaustive, but rather, is indicative of the quantity of literature pertaining to the "information explosion" through the years. It has appeared in a normal distribution frequency pattern, with most articles being published between 1965 and 1969. (This paper is my contribution to that same paper explosion.)

- Adkinson, W. A. and Stearns, C. M. "Libraries and Machines—A Review." *American Documentation* 18:121-4 (1967).
- Bar-Hillel, Yehoshua. "Is Information Retrieval Approaching a Crisis?," in his *Language and Information*, p.365-72. Reading, Mass.: Addison-Wesley, 1964. (Also in *Amer. Doc.* 14:95-8 (1963).)
- Beltran, A. A. "Evaluation of DDC TAB from Various Points of View," in Regional Workshop on the Report Literature, Albuquerque, N.M., *Proceedings*, p.78-85. Western Periodicals, 1965.
- Bernal, J. D. *The Social Function of Science*. London: Routledge, 1939.
- . "The Supply of Information to the Scientist: Some Problems of the Present Day." *Journal of Documentation* 13:195-208 (1957).
- . "Transmission of Scientific Information: A User's Analysis," in International Conference on Scientific Information, Washington, D.C., *Proceedings*, v.1, p.77-95. Washington, D.C.: National Academy of Sciences, Nat'l Research Council, 1959.
- Bourne, C. P. *Methods of Information Handling*. N.Y., Wiley, 1963. p.1-10.
- . "The World's Technical Journal Literature: An Estimate of Volume, Origin, Language, Field, Indexing and Abstracting." *American Documentation* 13:159-68 (1962).
- Brown, L. "Opening Session Address," in International Conference on Scientific Information, Wash., D.C. *Proceedings*, v.1, p.3-8. Washington, D.C.: National Academy of Sciences, Nat'l Res. Council, 1959.
- Conference on Problems of Centralized

- Documentation. *Proceedings*. . . . June 1949. (ATI 73852) p.37.
- East Coast Navy Interlaboratory Committee on Editing and Publishing. *Some Views on Dual Publication*, by F. H. Hunt. 1 Oct. 1968. (Monograph no. 5; AD676523) 5p.
- Ebersole, J. L. "An Operating Model of a National Information System." *American Documentation* 17:33-40 (1966).
- Federal Council for Science and Technology. *The Role of the Technical Report in Scientific and Technological Communication*. Dec. 1968. (PB180944) 108p.
- Freeman, M. E. "The Science Information Exchange as a Source of Information." *Special Libraries* 59:86 (1968).
- Glazier, K. M. "Paper Pollution." *Wilson Library Bulletin* 44:856-7 (1970).
- Goudsmit, S. A. "Is the Literature Worth Retrieving?" *Physics Today* 19:52-5 (1966).
- Gray, D. E. and S. Rosenborg. "Do Technical Reports Become Published Papers?" *Physics Today* 10(6):18-21 (June 1957). Also published as: Library of Congress. *Final Report on Study of Publication Stemming from Defense-Related Technical Reports*, Dec. 1956.
- Heumann, Karl. "Report Utilization Through Intelligence Sections," in B. M. Fry, and J. J. Kortendick, eds., *The Production and Use of Technical Reports*, p.111-15. Washington, D.C.: Catholic University, 1955.
- "The Information Deluge." *The Johns Hopkins Magazine* 17(4):1-33 (Fall 1967).
- Johns Hopkins University. *Reports of Studies of the Publication Fate of Material Presented at National Meetings*, June 1969. (TN 10, TN 12. PB185 469).
- Kent, A. "Resolution of the Literature Crisis in the Decade 1961-1970." *Research Management* 5:49-58 (1962).
- Lingeman, Richard. "American Notebook." *New York Times Book Review* (Mar. 1, 1970), p.24.
- London, Gertrude. "The Publication Inflation." *American Documentation* 19:137-41 (1968).
- Mantell, L. H. "On Laws of Special Abilities and the Production of Scientific Literature." *American Documentation* 17:8-16 (1966).
- Mellanby, K. A. "A Damp Squib." *New Scientist* 33:626-72 (1967).
- Miller, E. E. "The Genesis and Characteristics of Report Literature." *American Documentation* 3:91-4 (1952).
- Moore, C. A. "Preprints: An Old Information Device with New Outlooks." *Journal of Chemical Documentation* 5:126-8 (1965).
- "The National Information Problem." *SDC Magazine* 9:1-15 (1966).
- National Science Foundation. *Preliminary Study of Unclassified Government Technical Reports*, 31 July 1952. 6p.
- Overhage, C. F. J. "Science Libraries: Prospects and Problems." *Science* 155:802-6 (1967).
- President's Science Advisory Committee. *Science, Government, and Information*. Washington, D.C.: Govt. Print. Off., 1963. p.19-20.
- Price, D. J. deS. "A Calculus of Science." *International Science and Technology*, no. 15, p.37-42 (March 1963).
- . "The Ethics of Scientific Publication." *Science* 144:655-7 (1964).
- . *Little Science, Big Science*. New York: Columbia, 1963.
- Scrivenor, Thomas. "The Growth of Scientific Literature," in ASLIB 38th Annual Conference, 1964. *Looking Forward in Documentation*. London, 1964. p.3-9.
- U.S., Congress. House. Select Committee on Government Operations, 88th, 2d sess. *Documentation and Dissemination of Research and Development Results*. Washington, D.C.: Govt. Print. Off., 1964. (Study no. 4) 48p.
- Urquhart, D. J. "Rising Tide of Paper." *Advancement of Science* 21:279-85 (1964).
- Wall, Eugene. "A Rationale for Attaching Information Problems." *American Documentation* 18:97-103 (1967).
- Waterman, A. T. "The Future of Report Literature," in B. M. Fry and J. J. Kortendick, eds., *The Production and Use of Technical Reports*, p.3-8. Washington, D. C.: Catholic University, 1955.
- Wilson, J. H. "As We May Have Thought," in American Documentation Institute, *Proceedings*, v.3 p.117-122 Adrienne, 1966.