

# The Development of a South African Computerised Bibliographic Data Base on Water

P.J. AUCAMP

[SOUTH AFRICAN INFORMATION CENTRE, P.O. BOX 395, PRETORIA, 0001]

## Abstract

Since 1966 the National Institute for Water Research of the South African Council for Scientific and Industrial Research has issued a current awareness bulletin with a Keyword-in-Context index. This index and bulletin have been transformed to a bibliographic data base, the first in the water field in South Africa. The results of an investigation which led to this transformation are presented and discussed. The most important of these results is the fact that discipline orientated data bases are inadequate to cover the needs of a mission orientated field such as water research.

## Introduction

It is virtually impossible for the modern scientist to keep abreast of the vast amount of literature published annually, but secondary literature sources alleviate this burden. Water research, however, is a mission orientated field, which is not well served with secondary literature (Hooper, 1974) in contrast to the discipline orientated fields, such as pure chemistry or physics. This situation led to the creation in 1966 of *Current Literature on Water (CLOW)*, a current awareness bulletin and a Keyword-in-Context index to serve the interests of the National Institute for Water Research (NIWR) of the Council for Scientific and Industrial Research (CSIR). New developments, however, led to its abandonment in 1975, in favour of a more sophisticated data base.

## Establishment of CLOW

In 1966 the National Institute for Water Research requested the Information and Research Services (IRS) (both of the CSIR) to develop a current awareness service. The following requirements were set:

- The service should be truly current and as up to date as possible
- The coverage should be as wide as possible and include the whole interest field of the NIWR
- Some means of information retrieval must be available

IRS investigated the matter and decided that a system based on the Keyword-in-Context (KWIC) concept would satisfy these needs. The KWIC method, previously described by Luhn (1960) has been used extensively for both current awareness and retrieval purposes.

The current awareness system known as *Current Literature on Water* was managed by IRS and has been described by Lodder (1968). The volume of available literature, and the personnel involved, had a marked influence on the number of items indexed for CLOW (Figure 1). In 1975 two scientists

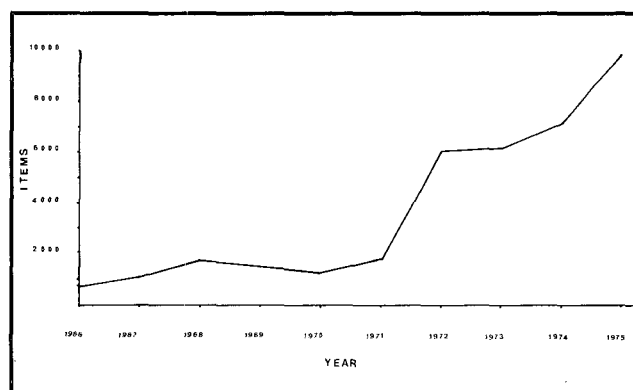


Figure 1  
Items in CLOW per annum

were responsible for indexing and for most of the administration. These indexers performed very well, each producing an average of about 5 000 items annually, compared with an average of 800 per abstractor as quoted by Martyn (1970). In addition, one punch machine operator was assigned to the punching of cards for the production of CLOW. The indexing policy stayed remarkably constant during the nine years of its existence. The most important changes are indicated in Table 1. The cost per item remained stable, although the total cost increased sharply (Table 2). The cost, however, compares favourably with indexing costs reported by similar organisations (Ashmole, 1973; Landau 1969; Nightingale 1973; Vickers 1973).

TABLE 1

## CHANGES IN THE INDEXING POLICY

1966 – July	First issue – weekly
November	Indication of language of the original
1976 – April	Fortnightly
1968 – March	Numbers allocated to pamphlets
May	Keywords allocated to some articles
	Patents included
	Monthly
1970 – March	Keywords allocated to all articles
1971 – October	Acquisition numbers allocated to documents
1973 – April	Type of article indicated
1974 – September	Temporary cutback in journals
1975 – February	Backlog wiped out
June	Administration transferred to the South African Water Information Centre
September	Last printed issue
December	Change over to the new bibliographic data base <i>Waterlit</i>

## Need for a National Water Information Service

In 1971 the Water Research Commission was established to promote and co-ordinate research in connection with water affairs in the Republic of South Africa. A definite need for a national information service in the water field was identified and the Commission therefore initiated the South African Water Information Centre in 1974. In June 1975 the Centre assumed responsibility for the administration of CLOW. It was required of the South African Water Information Centre to provide information for the whole water resources community in the country, including hydrologists, engineers and agriculturists, and CLOW therefore had to be adapted to these new needs.

In the meantime SASDI (Brunt 1974; Spicer 1974; Strumpher 1975), a computerized selective dissemination of information service, had been established at the CSIR and has been in operation since 1973. SASDI subscribes to the follow-

ing magnetic tapes: *Chemical Abstracts Condensates*, *Biosis Previews*, *Compendex*, *Inspec* and *Science Citation Index*. This service may be extended in the near future to include the following tapes: *Commonwealth Agricultural Bureau*, *Psychological Abstracts* and *Georef*.

The Centre, however, identified overlapping between SASDI and CLOW which justified investigation. About 250 journals were regularly scanned for CLOW. Most of the information on a specific subject comes from a nucleus consisting of a few periodicals (Goom 1974; Lancaster 1971) and in the case of CLOW only 50 journals (20%) yielded 50% of the indexed articles during the period 1971-1974. A list of the 25 most cited periodicals in CLOW appears in Table 3.

Further investigation revealed that 91% of the periodicals, screened by CLOW, were covered by at least one of the other commercially available abstracting services (*Engineering Index*, *Chemical Abstracts*, *Biological Abstracts* and *Inspec*) – See Figure 2. Five percent of these periodicals are covered by all four sources, 51% by at least three, 22% by at least two and 13% by one of the mentioned sources. Superficially this indicated a large duplication. However, when an attempt was made to trace the 707 periodical articles from 77 journals which were included in the June 1974 issue of CLOW in the above mentioned sources, the picture altered. Of these articles 20% appeared in *Chemical Abstracts*, 20% in *Biological Abstracts*, 13% in *Engineering Index* and 5% in *Selected Water Resources Abstracts*. Due to a certain amount of overlap between the sources a total of 42% was covered by one or more of these sources. This indicates that the discipline orientated services did not cover the mission orientated articles sufficiently and that the existence of CLOW or a similar service is justified. Blick and Magrill (1975) found a similar situation in the pharmaceutical industry, notwithstanding the fact that this industry is served by well developed bibliographic data bases such as *Drugdoc* and *Chemical Abstracts*.

Although the indexing for CLOW is done after the journal concerned has arrived in South Africa, which may sometimes be several months after publication, a comparison of the up-to-dateness of CLOW with that of *Chemical Abstracts*, *Biological Abstracts* or *Engineering Index*, has revealed that CLOW contains references to articles before most of the other sources (Table 4).

TABLE 2

## COST OF CLOW\*

	1967	1971	1975
<b>A. Printing</b>			
Monthly bulletin	780	—	4 800
Index	163	—	10 500
<b>TOTAL</b>	<b>R943**</b>	<b>R2 765</b>	<b>R15 300</b>
<b>B. Staff</b>	R2 400	R3 600	R15 540
<b>C. Total</b>	R3 343	R6 365	R30 840
<b>D. Items indexed</b>	1 092	1 775	9 812
<b>E. Cost/Item</b>	R3,06	R3,58	R3,14

\*Computer time and overhead costs are not included.

\*\*R1 = US \$1,15 (approximately)

TABLE 3

## LIST OF THE MOST CITED PERIODICALS IN CLOW ARRANGED IN DESCENDING ORDER

Periodicals	Times Cited in the Period 1971-1974
Journal of the Water Pollution Control Federation	587
Analytical Chemistry	458
Ground Water	346
RSA Patent Journal	340
Chemical Engineering Process	303
Water Research	301
Journal of the Fisheries Research Board of Canada	276
Journal of the American Water Works Association	262
Hydrobiology Journal	250
Water and Sewage Works	247
Lancet	243
Water Resources Research	226
Water Resources Bulletin	220
Water and Wastes Engineering	216
Applied Microbiology	216
Limnology and Oceanography	212
Analytical Chemica Acta	202
Water Pollution Control	186
Marine Biology	183
Archiv für Hydrobiologie	180
Hydrobiologica	165
Nature	138
Vom Wasser	137
Science	137
Desalination	127

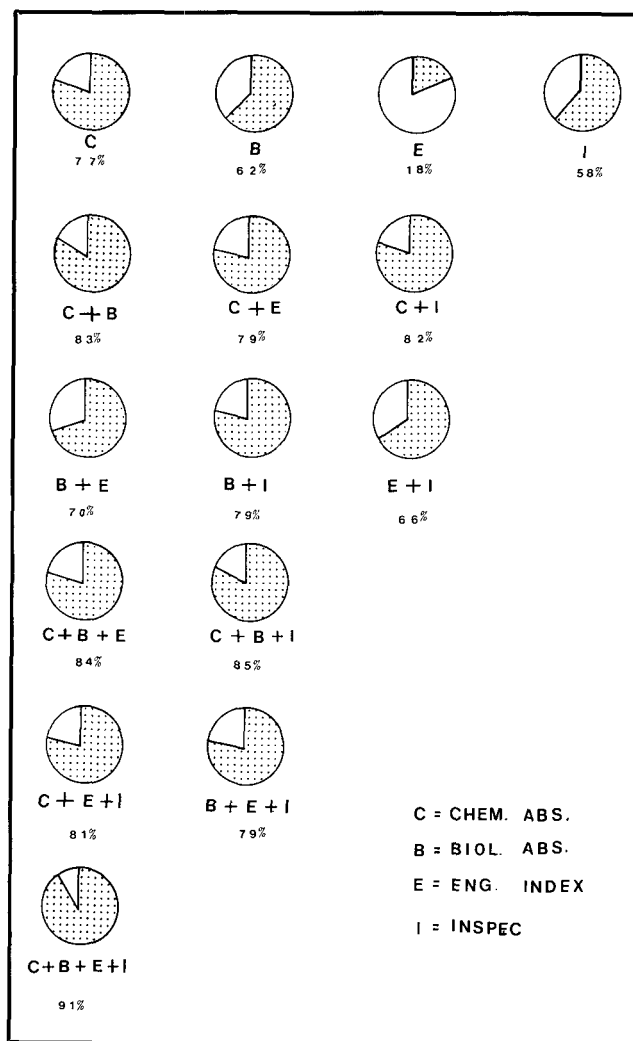


Figure 2  
Theoretical coverage of the periodicals in CLOW by combinations of other sources

CLOW contained many items not usually included in other secondary sources. The contents can be divided into patents (3%), short notes (10%), reports (27%) and periodical articles (60%). The KWIC index to CLOW also provides an index to the collection of reports, reprints and pamphlets held by the NIWR. As such it is an invaluable retrieval aid.

As an experiment all the CLOW data since 1971 were transferred to a magnetic disc which is now available via the Storage and Information Retrieval System (STAIRS) programme of IBM for interactive retrospective searching. An example of the results of a search is given in Table 5.

### Development of a New Bibliographic Data Base, WATERLIT

The results of the abovementioned investigations led to the decision to create a new bibliographic data base, WATERLIT, based on the old CLOW, but with a much wider subject division to include the interests of all the possible users in the water

TABLE 4

## UP-TO-DATENESS OF CLOW\*

Month	Commercially Available Abstracting Journals	CLOW
July	9%	31%
August	28%	31%
September	18%	13%
October	21%	13%
November	18%	6%
December	6%	6%

\*All the articles published in the June 1974 issue of 10 periodicals and that have been indexed/abstracted in CLOW has been traced in the other abstract journals. The percentage of the articles which appear in the various sources in the subsequent six months are indicated in the table.

TABLE 5

### EXAMPLE OF RETROSPECTIVE SEARCH USING THE CLOW DATA BASE

**Search – Query**

00001 Denitrification

**Doc. No.** JL 0004048**Title** Nitrogen removal from municipal waste water by columnar denitrification**Authors** Smith JM, Masse AN, Feige WA, Kamphake LJ.**Reference** Envir Sci Technol 1972 Vol 6 No 3 pp 260–267**Subj. Code** 0004**Keywords** Biological denitrification; methanol; biochemical reduction; nitrate.

TABLE 6

### CLASSIFICATION OF WATERLIT DATA

- 0000 **General**
- 0100 **Nature of Water**
- 0200 **Hydrological Cycle** (atmospheric water, catchments, runoff, soil water, groundwater, evaporation, transpiration and interception)
- 0300 **Resources**
- 0400 **Dams** (design, foundations, construction, outlet works, gates and spillways)
- 0500 **Aqueducts** (pipelines, canals and tunnels, pumps, flow measurement, gates and valves)
- 0600 **Conventional water supply** (reservoirs, wells and boreholes, reticulation systems, treatment processes)
- 0700 **Desalination** (processes, economics)
- 0800 **Water utilisation** (agricultural, forestry, domestic, mining and industry, recreational use, fish farming)
- 0900 **Water re-use** (agricultural, industrial, domestic, recreational)
- 1000 **Waste water** (domestic effluent, industrial effluent, mining effluent, agricultural effluent, storm water, ultimate disposal of waste, treatment for re-use, treatment processes, sewerage)
- 1100 **Solid waste**
- 1200 **Pollution** (sources of pollution and fate of pollutants, surface water pollution, ground water pollution, marine pollution, eutrophication, mineralization, legal control)
- 1300 **Quality analytical techniques** (chemical/physical, monitoring, biological, microbiological, sampling, instrumentation, standardization)
- 1400 **Limnology** (microbiological, botanical, zoological, chemical, physical)
- 1500 **Quality, criteria, standards and health aspects** (chemical, physical, biological, microbiological, surveys)
- 1600 **Planning and management** (data management, mathematical models, evaluation, project economics, financing, water demand and consumption patterns, management techniques, education and training, legal aspects, institutions, operation and maintenance, technology transfer, publicity, specialised information centres)
- 1700 **Marine aspects**

field in South Africa (Table 6). WATERLIT must be compatible with the computerized SDI service, SASDI of the CSIR, and must also be available for computerized retrospective literature searches via IBM's STAIRS programme.

A thesaurus has been compiled and rules have been laid down for the indexing of WATERLIT. This indexing will be done by the Centre as well as by other organisations and individuals. It is hoped that in this way a useful mission orientated bibliographic data base in the water field will be created.

## Conclusion

The experience gained with the production of CLOW has shown that a current awareness bulletin can be an invaluable aid to the scientist provided the volume of information is kept within limits. WATERLIT will use the SDI principle based on an interest profile for every scientist to ensure that he receives only that information in which he is interested. Whenever the need for a retrospective search in any field arises, the information will be available on-line. This will obviate the need for a KWIC index. The accent in WATERLIT will be on availability rather than on absolute comprehensiveness.

## REFERENCES

- ASHMOLE, R.F.; SMITH, D.E.; STERN, B.T.: Cost effectiveness of current awareness sources in the pharmaceutical industry. *Journal of the American Society for Information Science*. 29-39, Jan/Feb 1973.
- BLICK, A.R. and MAGRILL, D.S.: The value of a weekly in-house current awareness bulletin serving pharmaceutical research scientists. *The Information Scientist* 9(1). 19-28, 1975.
- BRUNT, A.G.: Ontwikkeling van 'n SDI-diens in Suidelike Afrika Deel I. *Suid-Afrikaanse Biblioteke* 41(4), 173-178, 1974.
- GOOM, H.H.: A computer based current awareness system producing both SDI output and conventional abstracts bulletin. *Aslib Proceedings*, 26(3) 98-108, 1974.
- HOOOPER, A.S.C.: Sources of information in water treatment research. *Water Research* 8, 101-107, 1974.
- LANCASTER, F.W.: The cost effectiveness analysis of information retrieval and dissemination systems. *Journal of the American Society for Information Sciences*. 12-27, Jan/Feb 1971.
- LANDAU, H.B.: The cost analysis of document surrogation – a Literature review. *American Documentation* 302-310, October 1969.
- LODDER, M.: A current awareness service based on the Keyword-in-Context (KWIC) indexing system. *South African Journal of Science* 64(5), 228-234, 1968.
- LUHN, H.P.: Keyword-in-Context index for technical literature (KWIC index). *American Documentation* 11, 228-295 (1960).
- MARTYN, J.: Evaluation of specialized information centres. *The Information Scientists*, 123-125, September 1970.
- NIGHTINGALE, R.A.: A cost benefit study of a manually-produced current awareness bulletin. *Aslib Proceedings* 25(4), 153-157, 1973.
- SPICER, N.: Development of an SDI service in Southern Africa Part 2. *South African Libraries* 41(4), 179-190, 1974.
- STRUMPFER, L. and BRUNT, A.G.: A SDI service in Southern Africa: SASDI Progress Report, March 1975. *South African Libraries* 42(5), 183-186, 1975.
- VICKERS, P.H.: A cost survey of mechanized information systems. *Journal of Documentation* 29(3), 258-280, 1973.