

Everything You Always Wanted To Know  
About ISI Data Bases  
But Were Afraid To Ask

Number 45, November 7, 1981

The variety of information services provided by the Institute for Scientific Information® sometimes leads to confusion. Because we offer both broad and specific information services for the sciences and social sciences, some people assume that we have several separate data bases. Others believe that ISI® has a single, monolithic data base and that we merely flip a switch to generate services for any disciplinary field.

Neither assumption is entirely correct. Whether we should be considered to have one data base or several depends on which service you are talking about and at what point in its production cycle it happens to be.

For example, the data bases for the *Science Citation Index*® and the *Social Sciences Citation Index*™ start out in our production cycle together, but eventually they split apart. *ASCA*® and *ASCATOPICS*® are always derived from the combined single data base. The data for the *Author Address Directory* and *Weekly Subject Index* sections of the six editions of *Current Contents*® are kept together through editing and keying procedures. An extraction program later replicates and separates the addresses and index terms for use in the different

*CC*® editions. This process is complicated by the fact that many journals covered in *CC* are not yet covered by one of our three citation indexes. (We have gradually been working towards the elimination of this confusing distinction.)

The raw materials for the data bases—the journals themselves—are not physically segregated by subject area during their processing flow through ISI. Social science journals lie contentedly next to physics and biology journals which are waiting to be processed. And the same is true in our library storage area. Our *Original Article Teletype Sheet (OATS*®) service makes no disciplinary distinctions among journals.

What is this all leading to? Recently I was invited to speak at the Third Institute of Electrical and Electronics Engineers (IEEE) Conference on Scientific Journals. Although I was unable to deliver the prepared paper in person, it is reprinted on the following pages. If, in the meantime, you have any questions frequently asked about ISI's operations. While there have been other articles about ISI as an organization,<sup>1,2</sup> this is the first one to describe our procedures in detail. While no single article of reasonable length can describe every

sentation of what actually hap-

the article does not include a dis-  
sion of the data base from which  
roduce *Current Abstracts of*  
*istry and Index Chemicus*<sup>™</sup>,  
*natic New Structure Alert*<sup>®</sup>,  
the *Index Chemicus Registry*  
*m*<sup>®</sup>. All of these services  
journal articles that announce  
organic syntheses and com-  
ds. For all intents and pur-  
s, all journals, editing, keying,  
ata processing for this data  
are handled separately from  
ISI services. Descriptions  
appeared in *CC* and  
here.<sup>3</sup>

roughout the article the enor-  
size of the ISI operation is  
clear. Over 2,000 new source  
are processed each day. This  
res over two million key  
es daily. I recently discussed in  
detail how we use our Key-  
system to cut costs and im-  
quality for this massive file.<sup>4</sup>  
her example of an innovative  
of the computer, not discussed  
the article, is our Project ZIP.  
ted simply, Project ZIP is de-

consistent or incomplete author ad-  
resses entered into the data base.  
Every organizational address pro-  
cessed is verified manually against  
an authority list of "correct" ad-  
resses. If the incoming address is  
not on the authority list, it is edited,  
assigned a number, and added to  
the list for future use. If an in-  
coming address is already on the  
authority list, the editor writes the  
appropriate code number beside it.  
When the article is processed, the  
data entry operator simply keys in  
the code number rather than the  
whole address. The correct and  
complete address is selected later  
from the computer's memory and  
automatically entered as part of the  
bibliographic record for that source  
item.

Like much published scientific re-  
search, the article reprinted here  
was not completely up to date even  
at the time it was presented. We are  
making improvements all the time.  
And next year, when we start our  
*Arts & Humanities Citation In-*  
*dex*<sup>5</sup> and our *Index to Scientific &*  
*Technical Proceedings*<sup>6</sup>, further  
refinements will be made.

---

**Garfield E.** The who and why of ISI. *Current Contents* No. 1, 6 January 1975,  
p. 5-14\*

**Lazerow S.** Institute for Scientific Information. *Encyclopedia of Library and*  
*Information Science*. Vol. 12. (A. Kent et al., eds.) New York: Marcel Dekker,  
1974, p. 89-97.

**Garfield E.** We've added a *Weekly Subject Index* to *Current Abstracts of*  
*Chemistry and Index Chemicus*. *Current Contents* No. 5, 3 February 1975,  
p. 5-6\*

-----, Project Keysave—ISI's new on-line system for keying citations cor-  
rects errors. *Current Contents* No. 7, 14 February 1977, p. 5-7.

-----, Will ISI's *Arts & Humanities Citation Index* revolutionize scholar-  
ship? *Current Contents* No. 32, 8 August 1977, p. 5-9.

-----, ISI's new *Index to Scientific & Technical Proceedings* lets you  
know what went on at a conference even if you stayed at home. *Current*  
*Contents* No. 40, 3 October 1977, p. 5-10.

inted in: **Garfield E.** *Essays of an information scientist*. Philadelphia: ISI Press, 1977.