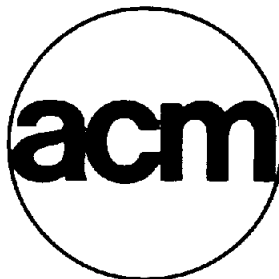




FDT ... BULLETIN OF ACM - SIGFIDET,  
THE SPECIAL INTEREST GROUP ON FILE  
DESCRIPTION AND TRANSLATION ....  
VOLUME 5, NUMBER 2 ... NOVEMBER, 1973

CONTENTS:

<i>SIGFIDET A Statement of Scope (Bernard Plagman)</i> .....	1
<i>MINUTES The SIGFIDET Annual Business Meeting</i> .....	2
<i>UNDERSTANDING RELATIONS (second installment) (E. F. Codd)</i> .....	5
<i>Summary of Past Codasyl DDLG Activity; Future Direction (E. Graf-Webster)</i> .....	7
<i>THE STREETWALKER FILE (Harold P. Sieglaff)</i> .....	9
<i>ASCII and EBCDIC Sequencing (John L. Little)</i> .....	15



F D T IS A PUBLICATION OF ACM'S SPECIAL INTEREST GROUP ON FILE DESCRIPTION AND TRANSLATION. IT IS PUBLISHED IRREGULARLY THREE TIMES A YEAR IN NEW YORK CITY. TECHNICAL PAPERS APPEARING IN THIS ISSUE ARE UNREFEREED WORKING PAPERS. ADDITIONAL COPIES AT \$2 EACH MAY BE OBTAINED FROM ACM HEADQUARTERS, 1133 AVENUE OF THE AMERICAS, NEW YORK, NY 10036 (TELEPHONE (212) 265-6300).

INQUIRIES CONCERNING SIGFIDET MEMBERSHIP SHOULD BE SENT EITHER TO ACM HEADQUARTERS OR TO SIGFIDET'S CHAIRMAN, BERNARD PLAGMAN, FEDERAL RESERVE BANK COMPUTER PLANNING DEPT, AJ, 33 LIBERTY ST., NEW YORK CITY 10045 (212) 732-5700, EXT. 8318.

*Material submitted for publication in FDT should be sent to the editor:*

*Randall Rustin  
Chase Manhattan Bank  
1 New York Plaza, 21st floor  
New York, N.Y. 10015*

*(212) 676-3831*

# SIGFIDET

## A Statement of Scope

Bernard Plagman

The current statement of scope for SIGFIDET was authored by the founder of FIDET, Don Hatfield, in 1969. Since that time, the technology of file description and translation has undergone notable change and the membership of SIG significantly broadened. These two developments seem to warrant a critical appraisal of the existing description of SIGFIDET activities.

Prior to any such considerations, it should be noted that a modification in the scope of a SIG is not a change of direction but an evolutionary process of changing emphasis, contributed to equally by the constituency of the SIG and its current leadership. Due to the dynamic nature of the technology we deal in, this is to be expected and moreover encouraged. Thus, a revised statement of scope could be viewed as an updating process by which the SIG maintains a current description of its special interests.

The single most significant development since the inception of SIGFIDET has been the emergence and acceptance of Data Base Management Systems. The activities within SIGFIDET, over the last few years, have been giving this area greater emphasis. Thus the following statement of scope is designed to emphasize this development:

Investigation of the areas of data management, data description, data structures and storage structures; examination of relationships between descriptor files or description languages and the organization of translators designed to use them to provide compatibility; creation of a forum on data description, data structures, storage structures and data management.

There are SIGS which have a defined scope which might raise issues with regard to overlapping scopes. SIG Information Retrieval (SIGIR), SIG Business Data Processing (SIGBDP) and SIG Programming Languages (SIGPLAN) are among these. In this regard, it would seem appropriate for the Chairman of these SIGs to consider the problem. The question has been raised concerning a change of the SIGFIDET name and acronym. Any SIGFIDETers with comments or suggestions concerning this should write to me directly.

### CURRENT STATEMENT OF SIGFIDET SCOPE

Investigation of problems of explicit description of data structures and formats; examination of relationships between descriptor files or description languages and the organization of translators which will use them to provide compatibility; creation of a forum on data and systems structures.

*Minutes*

The SIGFIDET Annual Business Meeting

August 28, 1973

The annual business meeting of SIGFIDET was held at ACM 73, Atlanta, Georgia, on Tuesday, August 28, 1973, at five thirty o'clock in the afternoon. Bernard Plagman, National Chairman, presided.

Board of Directors Present:

Randall Rustin	Newsletter Editor
Robert Lindner	Secretary-Treasurer
Kendall Wright	Past Chairman
Charles Bachman	
Eugene Raichelson	(External Activities Liaison)
Susan Brewer	(Standards Liaison)
Don Moehrke	

comprising a quorum of the Board.

Board of Directors Absent:

Edgar F. Codd  
Dennis Manelski  
T. William Olle  
Gene Altshuler  
Jim Fry

Chairman's Opening Remarks

The chairman discussed plans for activities during the next year. These included:

Newsletters (to be published 9/73, 12/73, 3/74, and 6/74)

Annual Workshop (Ann Arbor, Michigan, May 1-3, 1974)

A new activity/effort is under way to sponsor relevant research projects. At present, there are three (3) proposed subjects for these projects:

- 1) Data Administrator (DA) Implementation
- 2) Data Dictionary
- 3) Data Manipulation Language (DML) Interface

Suggestions for additional project subjects will be accepted for consideration by the Board. Contact B. Plagman.

Introduction of New Officers and Board Members

The chairman introduced the newly appointed officers and members of the Board of Directors. These are:

## *Minutes*

Susan Brewer - Board Member (and Standards Liaison)  
Jim Fry - Board Member  
Gene Altshuler - Board Member  
Robert Lindner - Secretary-Treasurer  
Randall Rustin - Newsletter Editor

### Planning for the Newsletter

The Newsletter Editor discussed plans for improving the newsletter in both content and format. In addition to this, the possibility of having SIGFIDET publish the newsletter instead of ACM is being studied. This would provide faster distribution of the newsletter. The feasibility of this proposal will require additional study from a budgetary point of view.

### The 1974 ACM-SIGFIDET Annual Workshop

The chairman announced the appointment of Gene Altshuler as General Chairman and Randall Rustin as Program Chairman for the annual workshop scheduled for May 1-3, 1974, in Ann Arbor, Michigan. Mr. Rustin announced the appointments of Jim Fry as Local Arrangements Chairman and Hal Uhrbach as one of the Session Chairmen. Plans are being made for six (6) sessions lasting three (3) days. The Program Chairman asked for topic suggestions for these sessions. He also stated that the Call for Papers is scheduled to be mailed at the end of September, and that those responding will be refereed. This statement raised the issue of having papers refereed as opposed to having the Program Committee review and select the papers. A discussion of this issue followed, resulting in a general consensus of opinion by those present that the papers submitted for the annual workshop should be reviewed and selected instead of refereed.

### Change of SIG Name

The chairman presented a proposal to change the FIDET name. To initiate discussion of this issue, the chair stated the following pros and cons:

#### Pros

- 1) The present name lacks identity (not clear)
- 2) The SIG's scope is not presently reflected by the term FIDET (File Description and Translation)

#### Cons

- 1) The effort will cost money:
  - a) to change existing publications
  - b) to hold a referendum as required by the SIGFIDET By-Laws
- 2) Some people readily identify with the present name

*Minutes*

- 3) Will take time and effort

A discussion followed concerning this issue. It was finally decided that for the present, SIGFIDET should formally but unofficially change its scope to more accurately reflect present activities. A request was made of the chair to write a formal statement outlining the present scope of SIGFIDET. The chair accepted this task. It was agreed that after this, if a name change is required, an appropriate one will be selected and a referendum will be held.

Financial Report

The chairman gave the financial report. SIGFIDET has an accumulated surplus of five thousand seven hundred and thirty dollars (\$5736) as of June 30, 1973. The Fiscal Year 1974 Budget has been submitted to ACM.

Membership

The chairman reported that as of June 30, 1973, SIGFIDET has eight hundred and ninety nine (899) members. This represents a twenty five per cent (25%) growth over the previous year. The chairman stated that the SIG/SIC Board is interested in expanding membership in all SIGs, and hopes to achieve this by generating a greater interest in ACM within the general business data processing community.

Questions

The chairman opened the meeting to questions. There were no questions from the floor.

Adjournment

The meeting was adjourned at six forty o'clock in the afternoon.

(Robert J. Lindner)  
Secretary-Treasurer

(Bernard K. Plagman)  
Chairman

## UNDERSTANDING RELATIONS

E. F. Codd

Q: Many people today are discussing the development of integrated data bases. Isn't an integrated data base merely a centralized pool of previously dispersed files?

A: Your question shows a confusion between centralization (a concept with geographical overtones) and integration (a non-geographic concept). The term integration applied to data, files, and data bases means bringing data together in such a way that its inherent semantic relationships can henceforth be exploited by a system (presumably a computer-based system). To accomplish this, a discipline needs to be imposed on the way data is represented, accessed, and modified. Bringing data together in this logical sense does not imply or require bringing it together in one physical location or even bringing it under the control of one human organizational unit.

Q: What is an "inherent semantic relationship"? Some examples might help.

A: In the first example a firm has developed separate and independent files concerning

- 1) employees and their skills.
- 2) projects and the employees assigned thereto.

Due to the separate development of these files (and of the application programs which interact with them) the derivation of what skills are being applied to any selected project(s) would entail a special programming effort to cross-relate employee data in the two files. This effort may include conversion of one or both files to a more common representation for employee identification. It is this employee identification which gives rise to one of possibly many inherent semantic relationships between the two files. Notice that in this example a degree of integration is possible without introducing new information.

In the second example, a firm T deals with customers through several branch locations. Each branch maintains an accounts receivable file and identifies its customers in its own fashion. Suppose that a customer at one branch, another customer at a second branch, and so on, are in fact themselves branches or subsidiaries of some single corporation S. The total indebtedness of S to T will, in these circumstances, be sufficiently difficult to ascertain that it will not be done on any regular basis. Customer identification in the branch files provides an inherent semantic relationship between them. Notice, however, that in this example fruitful integration necessitates the introduction of wholly new information: namely, which customers at which branches are members of a single larger institution.

## UNDERSTANDING RELATIONS

In both examples the inter-relatedness of data is the key ingredient. Exploiting this inter-relatedness is what integration is all about.

Q: Why is integration of data such a hot topic?

A: The primary motivation for integration (at least as far as commercial and industrial institutions are concerned) appears to be better management control of their operations. The potential reduction in redundancy of stored data appears to be more of an incidental pay-off than a motivation.

Q: If two previously logically separate files are to be inter-related in a machine-exploitable way, isn't it necessary to introduce pointers to link selected records in one file to related records in the other file?

A: No. The introduction of such pointers is only one way of solving the problem. We can look at this problem as one of defining a new relation (new to the system, that is) on already-established data. Although there are infinitely many solutions to this problem, only a few are practical from the standpoint of economy in storage space, access time or modification time. One such solution entails the introduction of a new file (which may be called a cross-index or linking file) consisting of records, each of which is itself a pair of record identifiers. One of these identifiers would uniquely identify a record in the first file, while the other identifier would uniquely identify a record in the second file -- more specifically, a record in the second file having the required relation to that in the first file.

It is not uncommon to find that the allegedly new relation that is desired is, in fact, derivable from the existing files through application of appropriate operators. In this case, the desired relation may be defined by an expression involving these operators, and no new data structure or data (other than the expression itself) need be added.

We have begun in this answer to touch upon an extremely important subject: a subject that is at the center of much of the controversy in the data base field today. The subject concerns the question: how can, and should, the data and relations in a formatted data base be represented to application programs and the many distinguishable varieties of end users? We shall attempt at least a clarification of this issue in later questions and answer.

Correspondence should be addressed to:

Dr. E. F. Codd, IBM Research Laboratory, K51/282, Monterey & Cottle Rds., San Jose, California 95193.

All rights reserved.

## SUMMARY OF PAST CODASYL DDLC ACTIVITY; FUTURE DIRECTION

### E. Graf-Webster

The CODASYL Data Description Language Committee was instituted on November 30, 1971, to take the work done by the CODASYL Data Base Task Group as a base and to develop from it a specification for a host-language independent data description language. The committee further formulated objectives to investigate related areas: the relation between the schema DDL and the host-language imbedded subschema DDL; the functional capabilities required in a Data Manipulation Language (DML) to properly interact with the schema; the provision of restructuring facilities in the DDL; and the provision of guidelines for subsetting of the DDL. The committee felt, however, that its most immediate purpose was the publication of a DDL specification, and that this was best accomplished by temporarily limiting its activities to clarification and only minor extensions of the DDL as specified in the DBTG '71 report. This purpose was accomplished as of the April 73 meeting, and a photo-ready copy of the new DDLC Journal of Development, containing the DDL specifications, has been sent to publishing organizations. Some of the differences between this document and the DBTG '71 report are as follows:

- (1) The distinction between the logical description of data structure and the physical implementation of the structure have been clarified. Physical connotations have generally been removed from this DDL.
- (2) The usage and format of data base procedures have been simplified and standardized.
- (3) The Set Selection Clause has been reformulated to make it more readable and easier to use.
- (4) Inconsistencies between the Set Selection and Location Mode Clauses have been eliminated.
- (5) The power of the CHECK clause has been considerably increased.

During the last few meetings, the committee has turned its attention to the longer range objectives stated above. The recent committee working papers indicate strong agreement among the members concerning the importance of certain topics to be considered by the committee. First among these are facilities which significantly enhance the DDL as, for example, the ability to do boolean set operations. Another topic is the separation and extension of the DDL into several sub-languages to encompass the various functional capabilities of pure logical structure description, strategic description of data usage, description of how the data is mapped out to physical devices, etc. Further considerations should also be given to data base restructuring capabilities, a subschema framework encompassing host-language independent subschema characteristics, and the generic DML capabilities.