ABSTRACT:
Medical institutions have difficulty tracking faculty productivity. To address this problem, the University of Nebraska College of Medicine has developed a computerized database containing faculty profiles, publications, education / teaching contact hours, and clinical service patient contacts. The database was designed in 1985 with the initial purpose of providing rapid and complete communication of research endeavors to university administrators and outside constituents.

SAS was selected at the time even though the project team had doubts that it was the best tool to handle this project. Lots of effort, and many versions of SAS code went into perfecting this system. Special direct access tables were built to reduce access time and improve I/O processing. These tables have proven to be very useful and efficient. This method is still far better than using SAS V6 indexes.

The system uses powerful but simple SAS techniques. This presentation shows samples of code, as well as a comparison between sequential and direct data access methods. The system runs on an IBM 9121 under VM/CMS and was developed using SAS, SAS/AF, and SAS/FSP.

FACULTY PRODUCTIVITY SYSTEM CONTENTS:
The system contains (1) faculty profiles - information about discipline and research specialization, as well as systems or diseases researched; (2) publications - faculty bibliographic information including papers, books and book chapters (abstracts and presentations are not included); (3) education / teaching activities - yearly estimates of teaching contact hours are collected from faculty representing lectures, clinical/small group teaching, and supervisory-advisory activities; (4) clinical service activities - data on patient contacts.

To make access to this information easier, a keyword thesaurus developed by the National Institutes of Health and the National Science Foundation (NIH/NSF) is used in coding all faculty profiles and publications. Additionally, a quantitative measure of publication productivity was developed using the Science Citation Index (SCI)'s impact factor of a journal (a ratio between citations and citable items published). Currently there are 4522 journals in the 1990 SCI. UNMC College of Medicine faculty have published in 1900 different journals since 1985.

DATA COLLECTION/UPDATE METHODS:
To maintain the accuracy and currency of this information, faculty activity printouts are generated and sent semi-annually. Faculty are asked to review the listings for accuracy and completeness. Changes are made on the printouts or by sending a photocopy of the first page of an article or by sending the relevant portions of their curriculum vitae. Quarterly UNMC library listings of faculty publications are also used to keep the database current and up-to-date. The data entry menu system was developed under SAS/AF and SAS/FSP.

To assure accurate and unique identification of faculty, author names follow a strict abbreviation scheme: Last name, followed by first initial and middle initial if any. In the case of multiple faculty ending up with the same author name, a distinction is tagged to each author name. A dictionary for non-UNMC faculty authors is maintained. Journal names follow a strict Index Medicus naming convention. SAS routines are used to catch erroneous entries and maintain accuracy and adherence to abbreviation and formatting standards.

PRODUCTIVITY ASSESSMENT PARAMETERS:
A SAS model using publications as the basic index of research productivity is used. Productivity assessments take into account the number of articles authored, byline position, number of authors for each article, and the citation Impact factor of the source journal. Alternative assessment algorithms were explored using different journal and byline position

909
weightings. Non-indexed journals received the minimum weight; books and book chapters received a weight based on selected assessment algorithm.

Faculty clinical service activities were tallied and summarized without differential weighting between the various patient contact categories. Faculty teaching contact hours were tallied and weighted based on the nature of the teaching activity.

**FACULTY RANKING:**
Faculty are ranked within their department and within the College of Medicine. Their productivity in publishing, teaching, and clinical service activities are compared to their respective department and college averages. These faculty productivity summaries are confidential; they are provided to the college dean's office and departmental chairmen. Faculty receive their own scores along with their department and college averages.

**PROGRAMMING METHODS:**
Several special considerations and techniques were needed in creating this information system. To collect publications into a SAS dataset, a structure may incorporate creating N author fields: author1, author2, ..., authorN. A problem with this structure exists for publications with more than N authors. Do you go back and add more fields, and so on infinitum? To make entry easier, the data entry screen had to look as similar as possible to formatting a curriculum vitae.

First Author: 
Co-Author(s): 

The first author is entered into its own field; Co-authors are entered in three multiple co-author fields: authors1, authors2, and authors3 (seventy five characters each). Co-authors are delimited by a comma, an author name can't be broken up between two fields. Using SAS functions such as: (substr, index, indexc, do while, etc.) author fields are parsed from the combined entry form to a singular form.

The original data entry set 'ENTRY.PAPER' is transformed to a temporary data set 'paper' containing only: Author name, and Publication number (NUMERO). Direct access could be used on this data structure, but since a faculty publications are not in physical sequence in the raw dataset, it would be necessary to use 'point' option and loop thru the datastep resulting in a terrible I/O performance.

```sas
DATA REPORT.PAPERS;
  MERGE PAPER(IN:IN1) ENTRY.PAPER(IN:IN2);
  BY NUMERO;
  IF IN1 AND IN2;
```

To solve this problem, the raw dataset is merged with the index file generating dataset 'REPORT.PAPERS', where a faculty publications are in physical sequence. It is larger than the initial dataset 'ENTRY.PAPER', but using direct access thru 'FIRSTOBS' and 'OBS' options, a monumental improvement in I/O performance occurs. It is the fastest method.

```sas
DATA PAPER (KEEP=AUTHOR BP EP);
  SET REPORT.PAPERS;
  BY AUTHOR NUMERO;
  LENGTH BP 4. EP 4.;
  RETAIN BP;
  IF FIRST.AUTHOR THEN BP= _N_; 
  IF LAST.AUTHOR THEN DO;
    EP= _N_;OUTPUT;END;
```

Now we have the following data structure: Author name, Beginning Publication observation number (BP), and Ending Publication observation number (EP).

Similar direct access tables are generated to access faculty profiles, education, clinical, NIH/NSF keyword codes, and faculty stats. All these tables were merged creating one global table containing pointers to all components.

**DESIGN OBJECTIVES:**
One design goal was to have one observation (containing all pointers) per faculty no matter how many components there are now or in the future, or how much data a faculty may have in each component. Another design goal was to avoid having costly IF statements and the like in generating reports. Last, but not least, it was important to create an easy to maintain SAS system environment, and a design method that can be prototyped throughout.

**REPORTS:**
Faculty activity printouts are generated and sent to faculty semi-annually (421 active faculty at this time). An activity report contains faculty profile, faculty productivity stats, bibliographic information, teaching contact hours, and clinical patient contacts. It would take terribly long to search 599 faculty profiles, 5433 publications, 2262 observations of education data, 1573 observations of summary patient clinical contacts, and 9709 observations of faculty yearly stats for one faculty, and then having to do this 421 times. By the end of the job 8,241,496 observations would have been accessed, excluding the cost of data step statements. It would be a terrible I/O performance...
job. Using a direct access method, a total of 18,961 observations would be accessed to do the whole job. The savings in CPU time is worth the effort of thinking and implementing such a scheme.

A sample portion of a faculty activity report code is as follows:

```sas
%MACRO ALL;
%LET P=0;
%DO I=1 TO &E;
DATA FINAL;
SET NAMES.UST (FIRSTOBS=&I OBS=&I)
/* direct access dataset */;
CALL SYMPUT ('PERSON',AUTHOR);
CALL SYMPUT ('F',F);
CALL SYMPUT ('BC',BC);
CALL SYMPUT ('EC',EC);
CALL SYMPUT ('BS',BS);
CALL SYMPUT ('ES',ES);
CALL SYMPUT ('BP',BP);
CALL SYMPUT ('EP',EP);
CALL SYMPUT ('BE',BE);
CALL SYMPUT ('EE',EE);
RUN;
*;
%LET PERSON = %QUOTE(&PERSON);
DATA NULL;
%LET P=0;
SET FINAL;
SET ENTRY.FACULTY (FIRSTOBS=&F OBS=&F);
FILE PRINT HEADER=H NOTITLES;
*** PRINT FACULTY PROFILE ***
RETURN;
H:P=SYMGET('P');P=P+1;
PUT @1 "&today" @55 P @SO "&PERSON";
CALL SYMPUT('P',P);RETURN;
RUN;
*;
%IF &BS NE. %THEN %DO;
DATA NULL ;
SET COM.STATS (FIRSTOBS=&BS OBS=&ES);
FILE PRINT HEADER=H NOTITLES;
*** PRINT FACULTY SUMMARY ***
RETURN;
H:...RETURN;
RUN;
%END;
*;
%IF &BP NE. %THEN %DO;
DATA NULL ;
FILE PRINT HEADER=H NOTITLES;
*** IF NO PUBLICATIONS ***
*** PRINT MESSAGE. ***
RETURN;
H:...RETURN;
%END;
*;
%ELSE %DO;
DATA NULL ;
SET REPORT.PAPERS (FIRSTOBS=&BP OBS=&EP);
FILE PRINT LINESLEFT=LL HEADER=H NOTITLES;
*** PRINT PUBLICATIONS ***
RETURN;
H:...RETURN;
%END;
%

For example, to print all the faculty in the department of Anatomy, set the following:
%LET B=1; /* First faculty to print */
%LET E=10; /* Last faculty to print */
Without a direct access method, this would be the total number of observations accessed (sequential access method): 599 (faculty profiles) + 5433 (publications) + 2262 (education) + 1573 (clinical) + 9709 (faculty stats) = 19576 (per faculty).

If a UNMC faculty member has no data, there is only a total of 2 observations (1 pointer/1 profile) accessed vs. 19576 observations (sequential access method).

INFORMATION PRESENTATION:
Ad-hoc reports are requested by departments for use in recruitment efforts and by college administration for review of research activities and resource allocation in specific subject areas. These reports include summaries at all levels of data (Table 1 - 8).

Many meaningful reports are created with little effort. For example:

**Faculty Journal Publishing Frequency**

<table>
<thead>
<tr>
<th>Faculty Name: Faculty X</th>
<th>Total no. of articles: xxx</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Journal(s)</strong></td>
<td>SCI Citation Impact Factor</td>
</tr>
<tr>
<td>Acta Cytol</td>
<td>1.149 0.899 0.845 0.847 1.030 0.970 xxx</td>
</tr>
<tr>
<td>Am J Clin Oncol</td>
<td>0.935 0.837 0.613 0.617 0.610 0.764 xxx</td>
</tr>
<tr>
<td>Am J Clin Pathol</td>
<td>2.449 1.971 1.971 1.969 2.081 1.858 xxx</td>
</tr>
</tbody>
</table>
CONCLUSION:
This system has proven to be a very useful productivity assessment resource, and provides us with the ability to quickly identify faculty and specific publications associated with any particular research area.

Using SAS, simple yet effective access methods make it possible to manipulate the data in an efficient manner.

REFERENCES:
7. Garfield, E. SCI Journal Citation Reports; a Bibliometric Analysis of Science Journals in the ISI Data Base.

Contact:
Mohamed Khachab
Senior Computer Analyst
Academic Computing Group
Department of Computing Services and Telecommunications
University of Nebraska Medical Center
600 South 42nd Street
Omaha, Ne. 68198-5030

Dan Moser, Ph.D.
Assistant Director
Academic Computing and Customer Support
Department of Computing Services and Telecommunications
University of Nebraska Medical Center
600 South 42nd Street
Omaha, Ne. 68198-5030

Robert Waldman, M.D.
Vice President for Medical Student and Graduate Education
AAMC
2450 N. Street, N.W.
Washington, DC 20037-1126

Leon B. Elwein, Ph.D.
Consultant to the Director
National Eye Institute
National Institutes of Health
Building 31, Room 6A06
Bethesda, Maryland 20892

SAS®, SAS/AF®, and SAS/FSP® are registered trademarks of SAS Institute Inc., Cary, NC, USA.
IBM® is a registered trademark of International Business Machines Corporation.
### Table 1: 
**Research Activity**  
<table>
<thead>
<tr>
<th>Year</th>
<th>Faculty X</th>
<th>Faculty Y</th>
<th>Non Weighted Rank</th>
<th>Weighted Rank</th>
<th>Non Weighted Rank</th>
<th>Weighted Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
</tr>
<tr>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
</tr>
</tbody>
</table>

### Table 2: 
**Clinical Service**  
<table>
<thead>
<tr>
<th>Year</th>
<th>Faculty X</th>
<th>Faculty Y</th>
<th>Non Heighted Rank</th>
<th>Heighted Rank</th>
<th>Non Heighted Rank</th>
<th>Heighted Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPT</td>
<td>DEPT</td>
<td>DEPT</td>
<td>DEPT</td>
<td>DEPT</td>
<td>DEPT</td>
<td>DEPT</td>
</tr>
<tr>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
</tr>
</tbody>
</table>

### Table 3: 
**Education Activity**  
<table>
<thead>
<tr>
<th>Year</th>
<th>Faculty X</th>
<th>Faculty Y</th>
<th>Non Heighted Rank</th>
<th>Heighted Rank</th>
<th>Non Heighted Rank</th>
<th>Heighted Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dept</td>
<td>Dept</td>
<td>Dept</td>
<td>Dept</td>
<td>Dept</td>
<td>Dept</td>
<td>Dept</td>
</tr>
<tr>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
</tr>
</tbody>
</table>

### Table 4: 
**Clinical Service**  
<table>
<thead>
<tr>
<th>Year</th>
<th>Faculty X</th>
<th>Faculty Y</th>
<th>Total</th>
<th>Dept</th>
<th>College</th>
<th>Inpatient</th>
<th>Outpatient</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
</tr>
</tbody>
</table>

### Table 5: 
**Research Activity**  
<table>
<thead>
<tr>
<th>Department</th>
<th>Year</th>
<th>Avg. # Faculty Per Year</th>
<th>Avg. # Clinicians Per Year</th>
<th>Avg. # Faculty Rank</th>
<th>Avg. # Clinicians Rank</th>
<th>Non Weighted Rank</th>
<th>Weighted Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
</tr>
<tr>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
</tr>
</tbody>
</table>

### Table 6: 
**No. of Publications**  
<table>
<thead>
<tr>
<th>Department</th>
<th>Year</th>
<th>No. of Publications Per Faculty-Year</th>
<th>Citation Impact Factor</th>
<th>No. of Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
</tr>
<tr>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
</tr>
</tbody>
</table>

### Table 7: 
**Education Activity**  
<table>
<thead>
<tr>
<th>Department</th>
<th>Year</th>
<th>Faculty</th>
<th>Small Group</th>
<th>Hours Rank</th>
<th>Hours Rank</th>
<th>Hours Rank</th>
<th>Hours Rank</th>
<th>Hours Rank</th>
<th>Hours Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
</tr>
<tr>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
</tr>
</tbody>
</table>

### Table 8: 
**Clinical Service**  
<table>
<thead>
<tr>
<th>Department</th>
<th>Year</th>
<th>Clinicians</th>
<th>Total Rank</th>
<th>Inpatient</th>
<th>Outpatient</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
<td>COLLEGE DEPT</td>
</tr>
<tr>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
</tr>
</tbody>
</table>

913