A Model for Database Design

Rebecca M. Renwick

Key Words: computers (general) • research

Computerized databases can facilitate several types of occupational therapy research. The value and usefulness of any database, however, is dependent on how well it has been designed. In this paper, a systematic, sequential-process model for the development of a computerized database is introduced. Each component of the model is illustrated by examples of its application to the actual design of a database for a community agency that provides occupational therapy services. The model focuses on issues related to the development of the contents of a database rather than on computer hardware and software. The issues addressed by the model include decisions about the purpose of the database, selection of the variables, and identification of the most appropriate measures with which to operationalize these variables. Content-related development issues have been given little attention in the literature, yet their neglect typically results in important limitations on the usefulness of a database. Therefore, this paper provides a set of guidelines for occupational therapists planning to establish a database for facilitating research.

A computerized, systematic body of information on clients is quite valuable to an occupational therapy department or agency. An advantage of this kind of database is its potential to facilitate one or more types of research, depending on the information it holds. For example, a computerized database could help clinicians answer research questions concerning the effectiveness of intervention programs (i.e., program evaluation and efficacy studies). Alternatively, it could facilitate the comparison of interventions based on different theoretical frameworks for a determination of the most effective approach for particular groups of clients.

The concept of a database is appealing, particularly because of recent advances in computer technology and software packages as well as the availability of a growing and detailed body of associated literature (Seymour, 1988). Although the establishment of an occupational therapy database might at first seem easy, as several authors have indicated, it is not quite as straightforward and free of pitfalls as one might expect (Brower, ten Kalten, & Meester, 1984; Renwick, 1989; Testa & Simonson, 1985; Thomsen, Laursen, Larsen, & Mellergard, 1981). In fact, the major difficulties encountered by the users of a database can usually be traced to flaws in the content design, that is, to the kinds of decisions made during the planning stages of database development (Renwick, 1989; Testa & Simonson, 1985; World Health Organization, 1977). Unfortunately, the resulting limitations in the usefulness of a database are often not recognized until long after the planning stages, that is, when its users attempt to draw on the information it incorporates. (See Brower et al., 1984, and Thomsen et al., 1981, for clear examples of this.)

Of course, the ultimate success or usefulness of a database also depends on the making of appropriate technical decisions. These sorts of decisions pertain to both hardware (e.g., computer equipment that will accommodate the volume of data to be input and analyzed) and software (e.g., programs that will manage and analyze the data effectively). Abundant generic literature is available on these technological issues. Therefore, although only a small portion of the literature deals specifically with rehabilitation databases (Renwick, 1989), this literature could easily be applied to the establishment of an occupational therapy database. It is important to note, however, that the literature on databases emphasizes these technical considerations to such an extent that, by implication, content design or nontechnical issues appear to be almost minor or peripheral. The focus of the present paper, therefore, is on these content design issues, that is, on issues that must be considered during the process of planning for the contents of the database. A sequential-process model for the designing of a computerized occupational therapy database is presented here. The model is equally applicable to mainframe computers and microcomputers. It draws on the sparse, widely scattered litera-

Rebecca M. Renwick, PhD, OT(C), is Assistant Professor, Department of Rehabilitation Medicine, and Associate, Centre for Health Promotion, University of Toronto, 256 McCaul Street, Toronto, Ontario, Canada M5T 1W5.

This article was accepted for publication April 26, 1991.
ture on these nontechnological considerations in the development of databases. Most of this literature concerns health care databases, because little information is available in the area of rehabilitation and virtually none exists on occupational therapy databases. The model also incorporates some innovative concepts and methods not previously applied to database development.

The sequential-process model presented here emphasizes a number of common issues inherent in the planning stages of database development. To the extent that these issues are not addressed adequately, the usefulness of the database will be limited. The model represents a framework that can help database developers avoid many of the planning stage pitfalls that can have unexpected and disconcerting repercussions for the database users. The five components of this model—identifying future uses for the database, selecting the overall approach to database development, identifying and selecting variables, identifying and selecting instrumentation, and testing instrumentation—are discussed and presented in the sequence in which they should be addressed. The rationale underlying each component is also discussed. Each model component is illustrated by examples of its application to a database project conducted at Community Occupational Therapy Associates, a large, community-based occupational therapy agency in Toronto, Ontario, Canada. This agency employs more than 100 occupational therapists and serves approximately 6,500 clients annually (Community Occupational Therapy Associates, 1990).

**Model Components**

**Identifying Future Uses for the Database**

Identification of the specific purposes of a database should be the first step in its development. When this process is neglected or its future uses are only vaguely defined, several negative consequences are likely to follow. One possible result, as noted by Brower et al. (1984), is that a great deal of tangential or irrelevant data is incorporated into the database on the premise that such data may be valuable to future users. Of course, this is costly in terms of the time that it takes staff to collect and enter the data into the database. Another possible consequence is that important information may be overlooked and, therefore, not incorporated (Thomsen et al., 1981). This also reduces the ultimate usefulness of the database. For example, the exclusion of such information may significantly restrict the kinds of questions that the database can be employed to answer.

At Community Occupational Therapy Associates, a small planning group of administrators and program managers with clinical responsibilities participated in the process of identifying the major purposes for their database. I was the project developer and leader. The group members identified several interrelated purposes for their database. The first was to facilitate the evaluation of Community Occupational Therapy Associates’ various client programs. This was considered vital, because ongoing program evaluation would provide information about the effectiveness of the agency’s intervention programs. Further, this kind of information would enhance the agency’s accountability to clients and third-party payers. The second purpose was to enable comparison of outcomes for clients in different functional groups who participated in the agency’s programs. The third purpose was to have the potential for future comparison of outcomes for groups of clients who shared the same functional problems but received different theory-based interventions. The fourth purpose was to be able to create profiles of the functional patterns of specific client groups before and after occupational therapy intervention.

**Selecting the Overall Approach to Database Development**

The second step of the model involves the selection of an overall approach to database development. A macro, or global, approach involves the creation of the database in the context of one complex, large-scale project (Sherman, 1981). A modular approach involves phases, or a series of smaller-scale projects (Booker & Platman, 1985; Wehr, 1987). The latter approach involves the development of one database segment or module at a time. It also typically requires a longer period of time to create the database, but it does allow the developers to learn as they go along. This can be a valuable process because developers’ experiences with early database modules can inform and facilitate the creation of later modules. A modular approach can also make the project less overwhelming: Database development in well-planned stages can make the overall project more manageable (Booker & Platman, 1985).

The group members chose a modular approach, which involves the creation of a module for one client group or program at a time. They decided to begin with a module on information concerning the psychosocial function of clients in their physical rehabilitation services, which is the agency’s largest program.

The clients in physical rehabilitation ranged in age from 16 to 80 years (88.0% of the clients were over 45 years of age, and 62.0% of the clients were women). Most (92.0%) lived in their own homes and did not live alone (67.5%). All of these clients had a primary diagnosis that was physical in nature. The most common diagnostic categories were neurological, musculoskeletal, and cardiorespiratory disorders. Many of these clients, however, also experienced some degree of psychosocial dysfunction (Community Occupational Therapy Associates, 1989).

This module would incorporate data on clients’ functional patterns before and after intervention as well as outcomes of intervention. The planning group members
were particularly interested in beginning with this module because they needed to have more complete and systematic data of this kind for the physical rehabilitation clients. The content of this module would partially address the overall purposes previously identified for the database as a whole. It should be noted that the same data could be used to answer questions about (a) intervention outcomes and (b) preintervention and postintervention functional patterns. The statistical procedures employed, however, would differ in each case.

The examples concerning application of the model presented in subsequent sections of the paper focus on this particular module on psychosocial functional and outcome measures. However, other modules incorporating different information on the physical rehabilitation clients (e.g., measures of physical function) or on other client groups (e.g., measures of psychosocial function in psychiatric clients) could be developed with the methods discussed in the following sections. Similarly, these methods could have been applied in the context of the more global approach to database development.

Identifying and Selecting Variables

The next step of the model involves identification of the potential variables to be incorporated in the module (or in the entire database, depending on the overall approach taken to development). At this stage of the content design process, it is advantageous for one to draw on the input of those who will collect and use the information in the database. Inclusion of these persons (or representatives of this group) in the development process can give them a sense of project ownership (Peters, 1986; Wehr, 1987) and, ultimately, promote enthusiasm about the database created. The literature on planned change within organizations clearly supports this method (see Kiresuk, Davis, & Lund, 1980, for a review). For the Community Occupational Therapy Associates project, this involved the addition to the original planning group of several occupational therapists who had considerable experience in physical medicine and psychosocial practice. These therapists were also likely to be among those who would gather the data and, eventually, use the information in this database module. This expanded planning group first identified the potential module variables and then selected the most valuable ones for inclusion in the module.

Important demographic and intervention-related variables (e.g., intervention goals, number of hours of intervention) are usually easy to identify. This was not a concern in the context of the current project, however, because the agency routinely collected detailed information of this kind on all its clients. Identification of the most important and meaningful information in other areas—in this case, which aspects of psychosocial function and outcome were most critical—is more difficult. A structured group approach can be used to ensure that only the most relevant variables are selected for inclusion. For this project, a nominal group process (Delbecq, Van de Ven, & Gustafson, 1975) was employed. Although this methodology has not previously been applied to the identification and selection of database variables, it is empirically based and highly efficient as compared with other group decision-making procedures (Delbecq et al., 1975). Because Delbecq et al. described this methodology in rich detail, only a brief description of the application of the process in the current context is presented here.

Each group member prepared for engagement in this decision-making process by reading theoretical literature, which I selected and distributed on important psychosocial issues affecting physical rehabilitation clients. The members also reflected on and discussed the clinical implications of these issues with their occupational therapy colleagues.

The planning group had decided in advance that a maximum of 10 variables relevant to client function and rehabilitation outcome would be included in this module. The purposes identified for the database made both preintervention and postintervention measures necessary. Further, the agency therapists would be collecting data with these measures. Thus, it was practical in terms of the therapists' time and costs to restrict this module to a maximum of 10 variables.

When the group members met to formally select the module variables, they reviewed their objectives for the session and the methodology for accomplishing them. Then each member independently generated what she considered to be the most important psychosocial measures of intervention outcome according to three criteria developed by the project leader, as follows: (a) high relevance to the intervention goals of most physical rehabilitation services clients, (b) strong likelihood of being affected directly or indirectly by occupational therapy intervention, and (c) high feasibility of being assessed by an occupational therapist in the clients' homes.

The group members generated 50 potential outcome measures that they then shared and discussed in terms of their meaning and conceptual basis. This was followed by a private, independent voting session. The voting procedure required each group member to select what she considered to be the 7 most important items (i.e., the 7 that best fulfilled the three criteria stated above) and then to rank them. The members' rankings for the items they had selected were pooled, and a list of the 20 items receiving the highest aggregate ratings was displayed for the group. Subsequently, the group reviewed and discussed in detail the meaning of these 20 potential variables. Each member then selected the 5 variables that best met the three original criteria and ranked these items. The results of this second vote were pooled, and the 10 items receiving the highest score were displayed for the group. These 10 items and their conceptual
Selection and Development of the Most Appropriate Instruments

The process of appraising potential instruments and measures in terms of their psychometric attributes and conceptual bases was carried out over several months by a research assistant with strong research skills. As the research assistant appraised various instruments, he also rated them according to the practicality criteria described above. The result was a short list of the two or three most psychometrically sound and practical measures that could be used to operationalize each of the 10 variables.

The planning group then considered this short list according to the three practicality criteria and the general acceptability of the wording and content of the items associated with each instrument. Each group member then independently ranked the instruments associated with each variable. There was strong agreement among group members on the ranking for all of the variables (see Table 2 for examples of some of these variables and the potential instrumentation associated with each).

As Table 2 indicates, no appropriate single measure of social support was available. Therefore, two existing instruments were selected to tap this construct. Further, no appropriate instrumentation was found to closely match the conceptual definition of motivation. Therefore, I developed a brief measure for this based on existing instruments and input from the project group. This process of developing new measures can be time-consuming but can result in instruments that are more accurate and, hence, yield more meaningful data.

Testing Instrumentation

The final step of the model, testing of the instrumentation selected with the client population of interest, can be

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Instrumentation for Operationalizing Module Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>Acceptable Instrumentation</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>Self-Esteem Scale (Rosenberg, 1965) Index of Self-Esteem (Hudson, 1982)</td>
</tr>
<tr>
<td>Control</td>
<td>Adult Nowicki-Strickland Internal-External Locus of Control Scale (Nowicki, 1978) Rotter’s Internal-External Locus of Control Scale (Rotter, 1966)</td>
</tr>
<tr>
<td>Social support</td>
<td>Social Support Questionnaire (Sarason, Sarason, Shearin, &amp; Pierce, 1987) Provisions of Social Relations Scale (Turner, Frankel, &amp; Levin, 1983) Social Interaction Dimension of Sickness Impact Profile (Gilson et al., 1975)</td>
</tr>
<tr>
<td>Motivation</td>
<td>No measures found to closely match conceptual definition — To be developed</td>
</tr>
<tr>
<td>Stress management/coping</td>
<td>Daily Stress Inventory (Brantley, Waggoner, Jones, &amp; Rapaport, 1987) Ways of Coping Checklist (Revised) (Vitiello, Russo, Carr, Mauro, &amp; Becker, 1985)</td>
</tr>
</tbody>
</table>

$^a$Selected to operationalize database variables.
done before or after actual programming for the database and data entry. The choice of timing in this regard will probably depend on the costs of programming and reprogramming, should the variables need to be refined, replaced, or excluded after testing. Whatever the decision concerning the timing of this test phase, it is an absolutely essential component of the model. The testing procedures employed should be as rigorous and careful as possible, to the extent that financial and human resources will allow (Renwick, 1989). This is vital, because the data collected and incorporated in the database will be meaningful and valuable only to the extent that the instrumentation for data collection is statistically reliable, valid, and sensitive to change for the particular client group. During the test phase, special attention should be given to difficulties with administration of the instrumentation as well as to problems experienced by clients responding to it (e.g., fatigue, difficulty with comprehension of items). On the basis of the results of this testing phase, both the instrumentation and the data-gathering procedures can be refined.

Some examples will illustrate how such refinements can be made. If clients’ overall (global) scores on two instruments are highly correlated, one of these instruments and the variable it measures may be excluded from the test battery and the database module. The same principle may be applied when there is a high correlation between two or more scales within an instrument or between particular scales of two different instruments; that is, one scale can be retained and the others can be omitted. Items that most subjects do not readily comprehend can be reworded and retested on a small sample. If during administration of the instrument battery, clients exhibit fatigue not relieved by brief rest periods, data collection could be done across two separate sessions. Alternatively, shortening of the test battery (based on the presence of high correlations between scores, as described above) may be an appropriate approach to reduction of clients’ fatigue.

At the time of this writing, the Community Occupational Therapy Associates project group was at the stage of being ready to test the instrumentation selected, prior to the programming for the database. This was a critical phase in the development of the database module because many of the psychosocial measures selected had not been previously used or widely used with persons with physical disabilities. Further, some of the instrumentation was developed specifically for this module. Thus, the psychometric properties of these measures needed to be either confirmed or determined for this client group.

Conclusion
Before the inception of this project, Community Occupational Therapy Associates already had in place a large-capacity mainframe computer as well as ongoing consultation from computer hardware and software specialists. Therefore, for the agency’s database project, technological issues could be addressed quite readily. Nevertheless, these technological issues are as important as content design. However, considerable literature (at various levels of complexity) exists on these generic technological issues. Additionally, numerous hardware and software consultants are available to database developers. In contrast, the literature on content design issues is sparse, and consultants who have the expertise to guide occupational therapy database developers through the content design process are rare. Thus, this paper focused exclusively on content-related issues.

Although the model presented was originally constructed for the development of a database for a community agency, it is versatile in its applicability. It can be used to guide the development of databases for other settings and client populations as well as for databases that facilitate other kinds of occupational therapy research. The model can usually be applied with few or no modifications. Employment of this model can have several major benefits. First, it can ensure that technological issues do not receive almost exclusive attention, to the detriment of the equally important content issues described here. Second, it will facilitate the process of programming for the database by permitting the database planners to specify the number, type, nature (e.g., level of data ratio, interval ratio), and format (e.g., actual column width) of each variable as well as the questions that the data will answer. Precise, accurate information of this kind gives the programmer and statistician clear guidelines, thus enabling them to develop programming that most appropriately and effectively addresses the needs of future database users. Third, it will greatly improve the likelihood of construction of a highly valuable and useful database. Although the application of this model requires a commitment of time and effort, it can facilitate the creation of a powerful research tool and a source of accurate, meaningful information about clients.

Acknowledgments
I gratefully acknowledge Community Occupational Therapy Associates of Toronto (Research and Program Development Fund) for financial support of this project.

This paper is based in part on presentations made at the annual conferences of the Canadian Psychological Association, Montreal, Quebec, Canada (June 1988), and the Canadian Association of Occupational Therapists, Winnipeg, Manitoba, Canada (June 1989).

References


**Coming in October:**

- The effect of reimbursement on occupational therapy practice
- A survey of physical modality use
- Play deprivation in children with physical disabilities
- The effects of task preference on performance
- Quality factors in research design

Turn to AJOT for the latest information on occupational therapy treatment modalities, aids and equipment, legal and social issues, education, and research.