Dagobert Soergel
College of Library and Information Services
University of Maryland
College Park, Maryland 20742
301-454-5453

How does one design a database?

Outline

Conceive idea

Agency databases - between public and personal (Figure 1).

Analyze requirements

Define user group. Internal vs. external.

Analyze users' background and tasks/problems and
resulting information needs (Figure 2).

Determine which of these information needs are or can
be met by existing databases.

Decide which of the remaining information needs should
be met through the new database (subject to an evolving
cost-benefit analysis).

Design the database

Decide on the coverage for the new database.

Prepare a list of potential data sources (Figure 3).

Develop the conceptual schema (Figure 4).

Decide on an approach to indexing (Figure 5).

Decide on an approach to searching and the user-system
interface (Figure 6).

Adapt or develop a thesaurus that supports the approach
selected.

Select tools for implementation (hardware, software, print)
and determine data structure/file formats.

Cost-benefit analysis in database design (Figure 7)

Problem- or request-orientation in database design (Figure 8)

Combining records extracted from multiple sources,
thus avoiding multiple database searches.

Including documents not found in public databases,
especially internal documents.

Searching a small database with a high concentration
of relevant documents.

Using one's own record format, index language and
perspective in indexing, geared to one's specific
interests.

Incorporating user's comments and notes (dynamic
database).

Efficient creation and upkeep of personal files.

Integration with other internal databases and other
functions, such as word processing.

Figure 1. **Reasons for a personal or agency
            database.**

Who are the people or organizations to be served (importance, backgrounds and skills, searching skills, physical location, means of communication)?

What are their tasks, decisions, problems?

What information is needed to solve these problems?

What needs are met now? How?

What searches are to be expected? (How many? How much variation? Requirements for answer quality? Deadlines?)

What is the present and expected impact of the information?

Figure 2. **Questions on users' backgrounds and needs.**

Public databases (but consider copyright).

      Regular searches, including SDI searches done for individual users.

      Retrospective searches done for individual users.

Library catalog.

Interlibrary loan records.

Catalog or other records of internal documents.

Personal files - distributed input.

Bibliographies of publications or other documents prepared by or for the agency.

Figure 3. **Sources for an internal database**.

Conceptual schema components

Record format. Data fields to be included

Display fields

Searchable fields (data structure)

Rules of form for data values

Rules for making entries

In design consider

Cost for data acquisition, input, and storage

Benefits in searching

Compatibility and linkage to other databases

Fields for user input and other detailed information

Design characteristics: Exhaustivity, specificity, flexibility, hospitality, compactness, efficiency, and reliability

Figure 4. **Conceptual schema.**

Request-oriented indexing, problem-oriented indexing

Index language is based on user's problems and requests expected. Index language controlled.

Indexer uses checklist of descriptors and makes relevance judgments.

Expensive. Performance good. Needed when searching for general and/or implied concepts.

Document-oriented indexing

Index language based on documents. Index language may be controlled or uncontrolled.

Indexer assigns descriptors to express the contents of the document. Alternatively, use words in the title, abstract, or full text.

Less expensive. Performance may be satisfactory if searches are for concepts that tend to be concrete and explicitly mentioned.

Exhaustivity and specificity of indexing

Figure 5. **Approaches to indexing**.

Type of search algorithm: Boolean, refinements for
free-text searching, ranked output.

Type of interaction with the system: Menu, script,
command language. Expert system for assistance to
the searcher. Relevance feedback. Only online or
also printed?

Thesaurus consultation online. May create dynamic
menus for descriptor selection.

Automatic use of thesaurus information in searching.
Inclusive searching (MEDLINE EXPLODE). Query
expansion in free-text searching.

Linkage to other databases and information
processing functions.

Entry of user comments to documents, both private
and public.

Figure 6. **Approach to searching.**
and user-system interface

Database building versus database use

Effort for indexing versus effort for searching.

Request-oriented versus document-oriented indexing.

Exhaustivity and specificity of indexing.

Free-text searching.

Cost for building and maintaining the data structure versus cost for searching.

Building indexes versus sequential searching.

Implementation of inclusive searching.

Difference in answer quality.

Search versus post-search screening

Figure 7. **Cost-benefit analysis as a design principle**.

Information system to support problem solving.

Active identification of users' problems and needs.

Need for substantive data rather than merely references. Chained search.

Request-oriented conceptual schema.

Request-oriented index language and indexing process.

Request-oriented data structure.

Figure 8. **Problem orientation as a design principle.**