

: What is the best approach for organizing data?

by Daniel J. Power

Editor, DSSResources.COM

In an information technology context, organizing means to arrange data in a coherent form and to systematize its retrieval and processing. When done well, organized data becomes an orderly, functional, structured collection. Today's reality is that data, including documents, come from many sources and can be used for multiple purposes. These data inputs need to be organized and stored for computer processing. How should we begin? What steps should we take to organize and reorganize data?

Data is an increasingly detailed reflection of reality. New data sources are becoming accessible and they need to be captured. Data streams, the sequence of digital data, can and does change. Data usage, how we use and process data, evolves. Historical data is not necessarily static and unchanging. Sometimes we reorganize data and sometimes we wish we could reorganize and supplement historical data and documents because of problems with the current data structure and organization. We organize data and documents in file folders by category, in data tables with links and relations, in flat files and in 3 dimensional column arrangements. Some data and documents have a natural hierarchical arrangement that is important to maintain. Some data has the possibility for many associations. Grouping, categorizing and labeling are all important skills associated with organizing data. Labeling is especially important and it involves describing identified data elements with words or short phrases. A good label adds meaning and helps in retrieval.

Organizing data and documents is a skill that can be learned and there are common organizing patterns used for similar organizing tasks. The three traditional database models or patterns are: 1) hierarchical model where data is organized into a tree-like structure, 2) network model that organizes data into records and sets, and 3) relational model where all data is represented in terms of tuples (an ordered list of elements/fields) grouped into relations. A database is an organized, coherent collection of data. The data is logically connected and consistent.

Data models differ in terms of specificity, perspective and scope. An enterprise data model provides a broad, abstract overview of all the data captured and stored in an entire organization. Kendle (2005) notes "An Enterprise Data Model is an integrated view of the data produced and consumed across an entire organization". In large organizations, constructing an enterprise data model can involve both simplification and aggregation. The high level model serves as a "means of visualization, as well as a framework supporting planning, building and implementation of data systems (Kendle, 2005)."

A conceptual project data model has a narrower, slightly more specific, yet abstract view of the data

: *What is the best approach for organizing data?*

relevant to the project. A project data model is a constrained subset of the Enterprise Data Model (cf., Friedgan, 1998). For a specific Information System project, multiple models are constructed that represent how data is logically organized and how it will be physically represented in files accessed by a computer's operating system.

Data models help visualize how groups of data called entities relate to one another. Data models can represent all sorts of data, from apartment-rental data to customer order data to zoo animal data. Multiple models can be constructed for a given data set to view the actual data ecology in a variety of ways and ensure that all relationships are represented. Just as ecology examines the relations of organisms to one another and to their physical surroundings, data ecology examines the relations of data to one another and to their organizational and computing surroundings.

The following steps can help organize data and documents:

1. Familiarize yourself with what needs to be organized. Examine forms and documents. If a data source or set has been previously organized, look at that structure and identify any problems encountered.
2. Try to identify patterns, clusters or categories in the data and documents. For a legacy paper form ask is some data repeated in multiple copies/submissions of the form? Is a data element associated with a particular day of the week or time of the year? Can you sort or group the data?
3. Keep a list of patterns and relationships. Look for multiple ways of organizing the data. Organizing is grouping like data together.
4. Define how you will use the data and documents. Decide if one of several organizing approaches fits better with your purpose than another.
5. Clearly define categories and put them in alphabetical or some sequential order. Look for overlaps in categories. Ask are the categories independent?
6. For structured data tables define fields. A field may link to a document and structured metadata is

: What is the best approach for organizing data?

stored in other fields in the record. For online files determine if sub files or nested files are helpful. Avoid complex file nesting. In general 3 levels should be sufficient.

7. Watch for data and documents that are in multiple files or tables. Do data or documents currently need to be copied or duplicated to improve retrieval? Can duplication and redundancy be avoided?

There is definitely some choice and discretion in organizing data and documents. One scheme(pronounced skeem), plan or design may be best for one purpose, but not for another very different purpose. One scheme may work well with a small amount of data, but not so well with large or very large amounts of data. Some organizing schemes work well when there are no updates or additions to a set of data and the data is static, but not so well with data changes and additions.

Who organizes data? The job title is varied, but "data analyst" is common. According to wiseGeek, "A data analyst is a highly trained individual that specializes in collecting, organizing, and analyzing data from various resources." Not all organizations have staff designated as data analysts, but the task of organizing data and documents must be done periodically and doing the task well is important.

There is **no** single "best" approach for organizing data. The steps and suggestions discussed are one approach that provides a general method for organizing data as electronic records. Maintaining an organized database can also be a challenge. New data and documents are added, there are changes in staff assigned to maintain the database, and there is an ongoing challenge with duplicated data and cleaning out obsolete, redundant and inaccurate data.

What is an electronic record? According to records.ncdcr.gov, "An electronic record is machine-readable, meaning that it requires hardware and software to be accessed and read. Organization is especially important so that these records can be found and retrieved. Electronic records include documents, spreadsheets, databases, images, video, and audio. If not managed, a computer assigns a unique name for these files when saved, but these names do not provide a context for the file, nor are they logical."

Organized data is potentially usable. Disorganized data is not. Electronic records must be organized and managed. To have value, data must be available and convenient for use.

: *What is the best approach for organizing data?*

References

Friedgan, A., "A Project Model is a Constrained Subset of an Enterprise Model," TDAN.com,, March 1, 1998 at URL <http://www.tdan.com/view-articles/4236>.

Kendle, N., "The Enterprise Data Model," TDAN.com, July 1, 2005 at URL <http://www.tdan.com/view-articles/5205> .

Microsoft, "File organization tips: 9 ideas for managing files and folders," at URL http://www.microsoft.com/atwork/productivity/files.aspx#fbid=MCDgf_LEUdy retrieved May 16, 2013.

North Carolina Department of Cultural Resources, "Best Practices for File-Naming," May 7, 2008 at URL http://www.records.ncdcr.gov/erecords/filenaming_20080508_final.pdf

<http://www.wisageek.com/what-does-a-data-analyst-do.htm>

Addendum

Part of organizing is naming and identifying what is stored. The North Carolina Department of Cultural Resources has guidelines online for naming files and electronic records:

Rule #1: Avoid using special characters in a file name. \ / : * ? " | [] & \$, .

Rule #2: Use underscores instead of periods or spaces.

Rule #3: Err on the side of brevity.

: *What is the best approach for organizing data?*

Rule #4: The file name should include all necessary descriptive information independent of where it is stored.

Rule #5: Include dates and format them consistently.

Rule #5: To more easily manage drafts and revisions, include a version number on these documents.

Rule #6: Be consistent.

Author: Daniel Power

Last update: 2013-07-25 05:11