Appendix I Data Management System Report
Current Data Management System for the Center for Writers

Introduction

Two purposes of a data management system (DMS) are to collect information for storage and to coordinate stored information for usage. With these two purposes as the focus, this project examines the existing DMS for the Center for Writers (the Center), a writing consultation service at North Dakota State University. The Center depends on the campus Information Technology Services (ITS) for most of its technology related needs. However, the Center has autonomy to select and implement a DMS. The system, therefore, reflects the information management decisions of the Center staff. A DMS was not foisted upon the Center. The examination of this project can concentrate on the DMS of the Center because a larger organizational assessment is unnecessary. The examination begins with an assessment of the DMS of the Center. The assessment functions as a basis for recommendations for the DMS of the Center. The scope of this report does not include budget projections or time tables because a needs assessment is the highest priority. After refining recommendations with the Center staff, proposals with budgets and time tables will be created.

Center for Writers Background

The services of the Center are available to all NDSU students and faculty. The Center employs two directors, four graduate consultants, and ten undergraduate consultants. The scope of the service is vast, as the Center accepts most written projects (essays, articles, dissertations, etc.) and serves a range of clients (first-year students to faculty). The diversity of the clients, therefore, requires attentive recordkeeping. The Center staff is technologically competent, and some of the staff members have technology training (MSOffice Suite or courses through ITS; programming experience with Java, PHP, Python, and others). No staff member is exclusively working on the technology needs. The tasks are shared among the staff members as needed.

Current Data Management System

The existing DMS is an Excel spreadsheet, with a direct entry method. The information originates from a manual filing system that includes three sources, two schedules (a graduate consultation schedule and an undergraduate consultation schedule) and a paper-record form. A fourth information source, a book and hand-out record, is being developed as another Excel sheet but that source is independent from the current DMS. Several staff members enter information into the schedules and paper-record, but one administrative assistant transcribes the data into the spreadsheet.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tutor</td>
<td>Day</td>
<td>Date</td>
<td>Hour</td>
<td>Class</td>
<td>Major</td>
<td>FirstName</td>
<td>LastName</td>
<td>Student ID</td>
<td>Dept</td>
<td>Course #</td>
<td>Instructor</td>
<td>Focus of Help</td>
<td>Gender</td>
</tr>
<tr>
<td>2</td>
<td>288</td>
<td>Kent TV</td>
<td>Wed</td>
<td>2 Feb</td>
<td>200 Sr</td>
<td>Zoology</td>
<td>Statement of Purpose</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The system has fourteen user set data types listed in Row 1 (Tutor, Day, Date, Hour, Class, Major, FirstName, LastName, Student ID, Dept, Course #, Instructor, Focus of Help, Gender). The input originates from a paper-record form that staff members complete when a client arrives, so value entry is contingent upon the staff obtaining the information from the client; the client does not enter data at any stage of collection. Several of the data type fields often are empty of values throughout the spreadsheet (Course # and Instructor).

The system is rudimentary. No arrays, base links, queue forms, query forms, embedded macros, pivot tables – or any other DMS feature exists. The spreadsheet is the only data table, so exporting and extracting the data involves typical Microsoft Office operations. The only query possible is the Excel built-in Find function. A manual document search is the only alternative query structure to the built-in query function. The query function of Excel (Sort) permits rearrangement of the values by selected columns within the table, however this function could create problems if the file is saved with a resort into chronological order (Date).

The restore point is similar to the system restore of the computer. Therefore the backup depends on the reliability of the machine, and the machine at present has experienced several virus attacks. The Center has a problem related to
data restoration because a pre-existing infection event is present. This problem means the computer system will experience a recurring registry error as the event by design spreads to unused file extensions, empty folders, and temp files. ComboFix, Malware, McAfee, and a manual Registry edit have failed to cleanse the system. The result is an unreliable electronic recovery system, so the paper-record forms are the understood as the backup. The paper-record forms were necessary as a backup when the Center lost data obtained using AccuTrack. AccuTrack experienced data collection problems, consequently most of the Fall 2010 data is only available in paper-record forms. Several email exchanges with AccuTrack support staff indicates the data is ‘simply lost’.

The purpose of the database is not inferential but descriptive. The descriptive target is the clients of the Center. The primary function of the accumulated descriptive data is to report client information for annual reports and other funding requests. A secondary function is to assist consultants by supplying information to prepare for return clients or, less frequently, supply information about clients for specific circumstances (such as informing an instructor of a client’s visit).

Assessment

The current DMS satisfies the primary function. A summation of client data is readily obtainable, and the annual reports can suffice with that descriptive data. However, the level of satisfaction is minimal, and the Center is actively seeking a new DMS. Additionally, inferential purposes could help the operation of the Center, and an electronic means of scheduling could further assist the Center through better organization and accessibility to information.

Efforts to parse the client’s by subject (data type: Class) provides little insight due to unidentified values in the paper-record or inconsistency from transcribing entries. For example, entered values for the Class ‘Polymers and Coatings’ include ‘Coatings’, ‘PC’, ‘CPM’, ‘Polymers’, and ‘Polymers and Coatings’. Those entries represent the most consistently misidentified values. Many other values of the class are identified with consistency. This error is a classic value definition error, where input incongruities result in ostensible DMS inefficiencies. Some subjects are overrepresented because of high frequency clients who have recurring appointments; other subjects are underrepresented because of value conflations. Some clients have as many as three majors during one semester, which is highly implausible.

One data type that causes systemic collection problems is the value ‘Exch’ that currently exists as a value within the data type ‘Class’. The intention of the value is to differentiate international students (or English language learners) from domestic students. The purpose of the differentiation is to know the frequencies of the two types of students. The tracking of the value experiences multiple breakdowns because the value should be a data type (two-values similar to data type ‘Gender’). The breakdown is attributable to the paper-record form that lists the ‘Exch’ value as a value for ‘Class’. The confusion of values at the data type level results in data entry errors. The errors accumulate into inaccurate claims about an important feature of the clients.

In addition to errors of collection, the Center has some errors of information gaps. The system does not have an obvious means of tracking medium of consultation sessions. If a data class such as ‘Online’ or ‘Computer’ existed, the information related to the means of conducting sessions would be available. Frequently, consultations involve email exchanges, working with a document on a student’s laptop, or downloading a file onto a Center computer. This class of information might increase because the Center offers WIMBA (online classroom) sessions and students might prefer working with non-print medium (the data does not exist on usage, so the argument is speculative only).

Some data types seem unnecessary only because a clear set of values does not exist or collection inconsistencies persist. The intent for collecting values associated with ‘Course #’ is to identify frequency of visitation by students of specific courses, which would allow the Center to prepare documents related to the course. The descriptive statistic has merit, but the inconsistent collection does not allow for further action at this time. The ‘Focus of Help’ data type needs some values assigned because the current collection method does not provide any insight. If the Center staff could identify trends in writing through a series of meetings, the services could have greater precision as consultants could anticipate client needs.

Concerns
The main concern is information passes through the system like water through a sieve. Strictly descriptive statistical data really only has one purpose, provide information for a report. Any inferences derived from the data are too speculative to structure a conclusion with significant confidence. It is an example of BFMI – brute force, massive ignorance. The Center hums with information as its service exists at a major node within the academic community of NDSU. Most majors of study and most levels of students will seek the service of the Center. But the information needs a conductor, that is, the information needs coordination through an efficient data management system. The Center already is an excellent writing assistance service, so fine tuning the DMS will serve to improve the service further in the aspects of information processing. The Center needs to identify the proper data classes, proper values, and explore methods to transform the cacophony into euphony.

Conclusion

The Center has a steady client base, but the data collection method contains flaws. The current data collection system requires some adjustment to explore additional and unaddressed variables, such as international student and type consultation type. The additional data provides information required assess whether the Center meets its stipulated mission to provide consultation services for all levels of writers. The additional data could allow the Center to assess its resource allocation to better address the needs of clients. Staff scheduling is important so consultants can help the greatest number of clients. The Center staff should develop a more systemic effort to collect information and prepare the staff to adhere to the system. A data management system is the first step in that development.