Understanding the Library’s Presence Online: Term Report

Stephanie Freas

Jingchen Sun

Alyson Youngblood

LIS 2680

March 29, 2013

**Problem Statement**

As libraries continue to grow their web presences, they often face difficult choices in how best to allocate funds and where to focus strategic growth efforts. Libraries need more information in order to make better decisions. A possible solution for this is to create a business intelligence database that stores information about libraries, their websites, and their web traffic, alongside other relevant information to assist libraries in benchmarking against peer organizations and making better, informed decisions regarding their web presence.

**Overview**

Although this database has the capacity to become an elaborate tool and information resource for libraries, its core focus is to store basic data:

* **about libraries**, such as name, type, and location;
* **about library websites**, such as URL, site size, and unique visitors;
* and **about library database subscriptions**.

The goal of this database is to provide libraries with a centralized location about the web presence of similar libraries for basic information needs. They will be able to access and retrieve information about libraries from their names, to their subscription count and details of their databases, to the size of their site, unique visitors, and the ways that the site has changed over time, based on the changes in visitor number, Facebook likes, and other information.

**Scope**

In reality, this database would likely exist behind a web frontend, likely for a business hoping to store basic information about libraries to subscribed users. However, for the purposes of this project, the database will exist on its own, but able to be expanded upon. Information from a total of six Pittsburgh-area libraries has been included in the database, from a variety of library types and sizes. These libraries include Carnegie Library of Pittsburgh, Carnegie Library of McKeesport, University of Pittsburgh Hillman Library, Carnegie Mellon University Hunt Library, University of Pittsburgh Health Sciences Library System, and the Byzantine Catholic Seminary Library. All information for this project was procured freely from the library websites.

**User Requirements**

* The database must store basic information about a library, its website, its database holdings, and its presence on the web, such as social media information, as well as other related information
* The database must be able to hold information about a multitude of libraries
* Information should be made available to subscribed users, for free or a low monthly fee
* The database must be able to hold information about all libraries in the Pittsburgh area
* Users must be able to quickly and efficiently run any queries
* The database must be created and run on Microsoft Access so that users may easily use it

**Conceptual Schema**

*Business Rules*

A library may have one website. A website must belong to only one library.

A website may have web traffic. Web traffic belongs to one website.

Web traffic must occur in a time period. A time period may contain traffic.

A website may include zero or more databases. A database may be available on zero or many websites.

A library may have zero or more blogs. A blog must belong to one library.

A library may have zero or more Twitter accounts. A Twitter account must belong to one library.

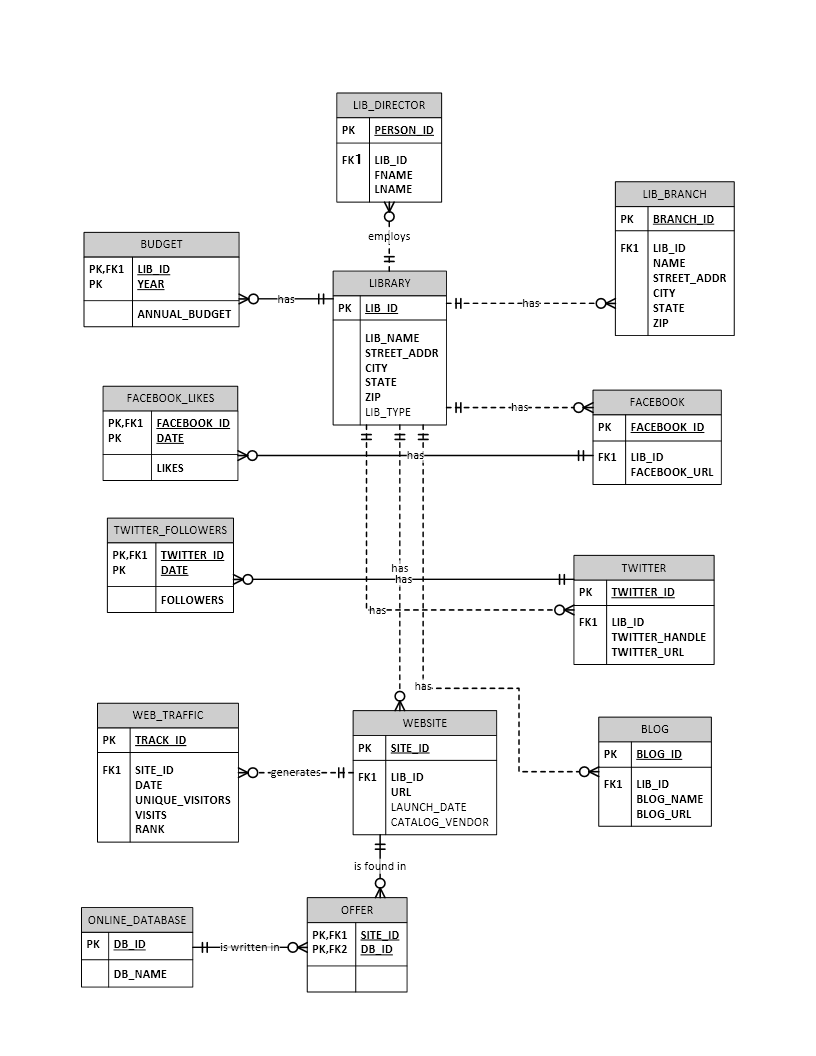
A library may have zero or more Facebook accounts. A Facebook account must belong to one library.

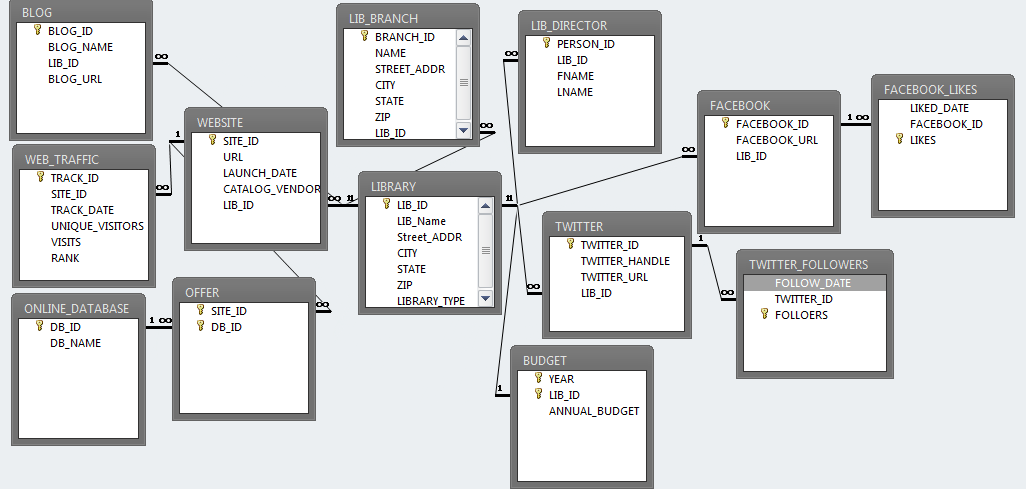
A library must have a budget. A budget must belong to one library.

A library may have one director. A director may belong to one or more libraries.

A library may have zero or more branches. A branch must belong to one library.

A library must have one or more types. A type may belong to one or more libraries.

***Figure 1.1 Conceptual Model of Database, created in Microsoft Visio*** 

**Logical Schema**

***Figure 1.2 Microsoft Access View of Entity Relationships in Database***

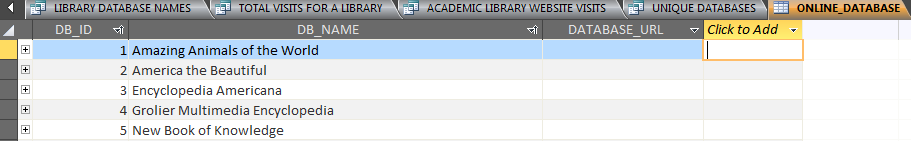
***Data Dictionary***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **TABLE NAME** | **ATTRIBUTE NAME** | **CONTENTS** | **TYPE** | **FORMAT** | **REQUIRED** | **PK or FK** | **FK REFERENCED TABLE** |
| LIBRARY | LIB\_ID | Library code | CHAR(10) | 9999999999 | Y | PK |  |
|  | LIB\_NAME | Library name | VARCHAR(100) | Xxxxxxxxxx | Y |  |  |
|  | STREET\_ADDR | Library street address | VARCHAR(50) | Xxxxxxxxxx | Y |  |  |
|  | CITY | Library city | VARCHAR(30) | Xxxxxxxxxx | Y |  |  |
|  | STATE | Library state | CHAR(2) | XX | Y |  |  |
|  | ZIP | Library ZIP code | CHAR(5) | 99999 | Y |  |  |
|  | LIB\_TYPE | Library type | VARCHAR(10) | Xxxxxxxxxx | Y |  |  |
| LIB\_DIRECTOR | PERSON\_ID | Library director code | CHAR(10) | 9999999999 | Y | PK |  |
|  | LIB\_ID | Library code | CHAR(10) | 9999999999 | Y | FK | LIBRARY |
|  | FNAME | Library director name | VARCHAR(20) | Xxxxxxxxxx | Y |  |  |
| LIB\_BRANCH | BRANCH\_ID | Branch code | CHAR(10) | 9999999999 | Y | PK |  |
|  | LIB\_ID | Library code | CHAR(10) | 9999999999 | Y | FK | LIBRARY |
|  | NAME | Branch name | VARCHAR(100) | Xxxxxxxxxx | Y |  |  |
|  | STREET\_ADDR | Branch street address | VARCHAR(50) | Xxxxxxxxxx | Y |  |  |
|  | CITY | Branch city | VARCHAR(30) | Xxxxxxxxxx | Y |  |  |
|  | ZIP | Branch zip | CHAR(5) | 99999 | Y |  |  |
| BUDGET | LIB\_ID | Library code | CHAR(10) | 9999999999 | Y | PK, FK | LIBRARY |
|  | YEAR | Budget year | CHAR(4) | 9999 | Y | PK |  |
|  | ANNUAL\_BUDGET | Annual budget | CURRENCY | $9,999,999.99 | Y |  |  |
| FACEBOOK | FACEBOOK\_ID | Facebook code | CHAR(10) | 9999999999 | Y | PK |  |
|  | LIB\_ID | Library code | CHAR(10) | 9999999999 | Y | FK | LIBRARY |
|  | FACEBOOK\_URL | Facebook URL | VARCHAR(50) | Xxxxxxxxxx | Y |  |  |
| TWITTER | TWITTER\_ID | Twitter code | CHAR(10) | 9999999999 | Y | PK |  |
|  | LIB\_ID | Library code | CHAR(10) | 9999999999 | Y | FK | LIBRARY |
|  | TWITTER\_HANDLE | Twitter handle | VARCHAR(30) | Xxxxxxxxxx | Y |  |  |
|  | TWITTER\_URL | Twitter URL | VARCHAR(40) | Xxxxxxxxxx | Y |  |  |
| TWITTER\_FOLLOWERS | TWITTER\_ID | Twitter code | CHAR(10) | Xxxxxxxxxx | Y | PK, FK | TWITTER |
|  | DATE | Followers date | DATETIME | dd-mm-yyyy | Y | PK |  |
|  | FOLLOWERS | Followers count | NUMERIC(10,0) | 9,999,999 | Y |  |  |
| FACEBOOK\_LIKES | FACEBOOK\_ID | Facebook code | CHAR(10) | 9999999999 | Y | PK, FK | FACEBOOK |
|  | DATE | Likes date | DATETIME | dd-mm-yyyy | Y | PK |  |
|  | LIKES | Likes count | NUMERIC(10,0) | 9,999,999 | Y |  |  |
| WEBSITE | SITE\_ID | Website code | CHAR(10) | 9999999999 | Y | PK |  |
|  | LIB\_ID | Library code | CHAR(10) | 9999999999 | Y | FK | LIBRARY |
|  | URL | Website URL | VARCHAR(40) | Xxxxxxxxxx | Y |  |  |
|  | LAUNCH\_DATE | Website launch date | DATETIME | dd-mm-yyyy |  |  |  |
|  | CATALOG\_VENDOR | Library catalog vendor | VARCHAR(30) | Xxxxxxxxxx |  |  |  |
| BLOG | BLOG\_ID | Blog code | CHAR(10) | 9999999999 | Y | PK |  |
|  | LIB\_ID | Library code | CHAR(10) | 9999999999 | Y | FK | LIBRARY |
|  | BLOG\_NAME | Blog name | VARCHAR(50) | Xxxxxxxxxx | Y |  |  |
|  | BLOG\_URL | Blog URL | VARCHAR(40) | Xxxxxxxxxx | Y |  |  |
| WEB\_TRAFFIC | TRACK\_ID | Tracking code | CHAR(10) | 9999999999 | Y | PK |  |
|  | SITE\_ID | Website code | CHAR(10) | 9999999999 | Y | FK | WEBSITE |
|  | DATE | Traffic date | DATETIME | dd-mm-yyyy | Y |  |  |
|  | UNIQUE\_VISITORS | Unique website visitors | NUMERIC(10,0) | 9,999,999 | Y |  |  |
|  | VISITS | Website visits | NUMERIC(10,0) | 9,999,999 | Y |  |  |
|  | RANK | Website rank | NUMERIC(10,0) | 9,999,999 |  |  |  |
| ONLINE\_DATABASE | DB\_ID | Database code | CHAR(10) | 9999999999 | Y | PK |  |
|  | DB\_NAME | Database name | VARCHAR(50) | Xxxxxxxxxx | Y |  |  |
| OFFER | SITE\_ID | Website code | VARCHAR(10) | Xxxxxxxxxx | Y | PK, FK | WEBSITE |
|  | DB\_ID | Database code | CHAR(10) | 9999999999 | Y | PK, FK | ONLINE\_DATABASE |

**Sample SQL Queries**

*Queries 1-3 are more appropriate for the database creators and changing the properties of the database rather than retrieving information. Queries 4-9 are used for retrieving specific data, something more helpful for users.*

1. Alter Auto-incrementing Number



ALTER TABLE ONLINE\_DATABASE

ALTER COLUMN DB\_ID AUTOINCREMENT(1,1);

*This query is meant to set the DB\_ID's next entity value as 1 and will increase one by one. It is useful for extensive lists of fields, so a person does not have to individually number every field and can use auto-generated values.*

1. Create A Table



CREATE TABLE OFFER (

SITE\_ID INT NOT NULL,

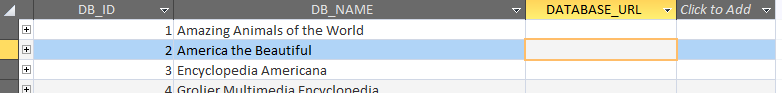
DB\_ID INT NOT NULL,

PRIMARY KEY (DB\_ID, SITE\_ID),

FOREIGN KEY (DB\_ID) REFERENCES ONLINE\_DATABASE,

FOREIGN KEY (SITE\_ID) REFERENCES WEBSITE);

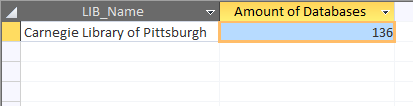
*This query is used to build a new table. Although this is possible without a SQL query, the creators wanted to utilize the SQL commands to become familiar with the tools.*

 3. Add a Column

ALTER TABLE ONLINE\_DATABASE ADD DATABASE\_URL TEXT;

*This query, ADD A COLUMN, is used to add an attribute to one table. When the query is executed, then a new attribute, DATABASE\_URL, with its defined data type (text) shows up in ONLINE\_DATABASE table. Although DATABASE\_URL was not added to our database, it is can be considered for future use.*

1. Library Database Count



SELECT L.LIB\_Name, COUNT (D.DB\_ID) AS [Amount of Databases]

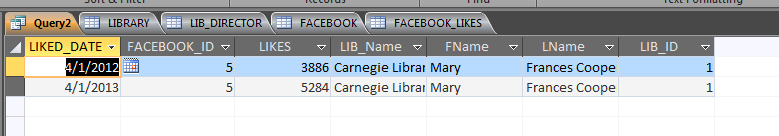
FROM LIBRARY AS L, ONLINE\_DATABASE AS D

WHERE D.DB\_ID IN (SELECT O.DB\_ID FROM OFFER AS O WHERE O.SITE\_ID=1)

AND L.LIB\_ID=1

GROUP BY L.LIB\_Name;

*This query was created to retrieve the name of a library and its amount of reference databases listed in our database. This example is Carnegie Library of Pittsburgh, which has a total of 136 databases. Libraries can use this information, alongside the size of the library (see library website visits or unique visitors) to see how many databases other libraries may have. While this query can be modified to show all libraries and their databases, this simple instance allows users to learn about a specific library holdings and benchmark against those.*

5. Facebook Likes Growth

SELECT LIKED\_DATE, A.FACEBOOK\_ID, LIKES, LIB\_Name, FName , LName, F.LIB\_ID

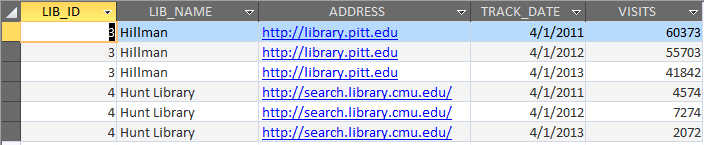
FROM FACEBOOK\_LIKES AS A, LIBRARY, LIB\_DIRECTOR, FACEBOOK AS F

WHERE LIB\_DIRECTOR.LIB\_ID=1 AND A.FACEBOOK\_ID=F.FACEBOOK\_ID AND F.LIB\_ID=LIBRARY.LIB\_ID AND LIBRARY.LIB\_ID=LIB\_DIRECTOR.LIB\_ID

GROUP BY LIKED\_DATE, A.FACEBOOK\_ID, LIKES, LIB\_Name, FName, LName, F.LIB\_ID;

*This query allows users to see the Facebook information for a particular library over time. By retrieving the “liked date” and the Facebook likes for Carnegie Library of Pittsburgh, users can see that the amount of Facebook likes grew from 3886 in April 2012 to 5284 in April 2013. Based on this information, libraries can make inferences about the online and Facebook presence of this library to consider for their own online presence. This may influence decisions about marketing or creating social media accounts.*

6. Academic Library Visits



SELECT LIBRARY.LIB\_ID, LIBRARY.LIB\_NAME, WEBSITE.URL, WEB\_TRAFFIC.TRACK\_DATE, WEB\_TRAFFIC.VISITS

FROM LIBRARY, WEBSITE, WEB\_TRAFFIC

WHERE LIBRARY\_TYPE= 'ACADEMIC' AND LIBRARY.LIB\_ID=WEBSITE.LIB\_ID AND WEBSITE.SITE\_ID=WEB\_TRAFFIC.SITE\_ID

GROUP BY LIBRARY.LIB\_NAME, LIBRARY.LIB\_ID, WEBSITE.URL, WEB\_TRAFFIC.TRACK\_DATE, WEB\_TRAFFIC.VISITS;

*This query retrieves the information for website visits every year for academic libraries. It is helpful to see how many times users of a specific academic library visit their webpage. With more information, such as when a library may add new resources or change the site, it will be helpful to consider how those changes impact the library.*

7. Library Database Names



SELECT DB\_NAME, LIBRARY.LIB\_Name

FROM ONLINE\_DATABASE, OFFER, WEBSITE, LIBRARY

WHERE ONLINE\_DATABASE.DB\_ID=OFFER.DB\_ID AND OFFER.SITE\_ID=WEBSITE.SITE\_ID AND WEBSITE.LIB\_ID=LIBRARY.LIB\_ID AND LIBRARY.LIB\_Name='Hunt Library';

*This query allows users to see a named list of the databases a library holds. In this instance, the information is listed for a specific library, Hunt Library. This query is useful when a library trying to purchase new databases would like to see what other libraries may have.*

8. Total Website Visitors for a Library



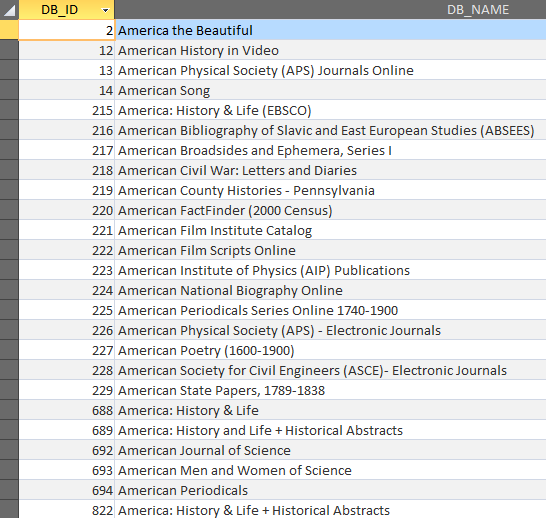
SELECT SUM (VISITS) AS [Total Visitors of BCS Library]

FROM WEB\_TRAFFIC, WEBSITE

WHERE WEB\_TRAFFIC.SITE\_ID=WEBSITE.SITE\_ID AND WEBSITE.LIB\_ID=6;

*This query quantifies the sum of all visitors for a library website. In this case, it is for the Byzantine Catholic Seminary Library, and users can see that since the data was initially collected, there have been a total of 1951 visits to the website. Users may use this to see the total number of visits for a specific library and compare it to other types of information, such as social media presence, database offerings, and budget. This query can also be a simple building block for more complex queries.*

9. Unique Databases



SELECT \*

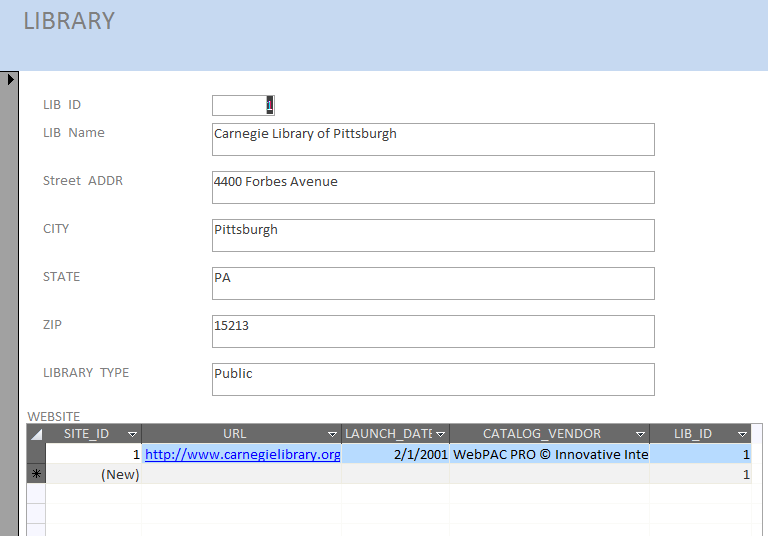
FROM ONLINE\_DATABASE

WHERE DB\_NAME

LIKE 'AMERICA\*';

*This query is also helpful for users who want to gather information about what different kinds of databases are available. The specific databases here have the word “America” in them. Unlike many other queries, this allows users to focus on something other than specific library information. This query is very helpful for those who want an expansive, detailed list of what databases are not only available, but also purchased and owned by other libraries.*

***Sample Form and Report***



\*LIBRARY Input Main Form\*

*This form allows users to input the specific information about a library, website, budget, and online databases. While there is a significant amount of more information about the libraries to input, this can be considered one part of the forms and be used to create others.*



Current Libraries Report

*This report is helpful for users who want to access the concise, yet important information about the libraries in the database. It includes the LIB\_ID to reference other tables, the name and type of the library, the URL, and when the URL was launched, for background information. Because the library is the central table and entity of this database, reports will mostly focus on the library information. This report is an example of what information an end-user might request.*

References and Helpful Websites

*Library websites we used to gather the information for our databases:*

Carnegie Library of Pittsburgh <http://www.carnegielibrary.org/>  
Carnegie Library of McKeesport <http://www.mckeesportlibrary.org/>

University of Pittsburgh Library System <http://library.pitt.edu>

Carnegie Mellon University Hunt Library <http://search.library.cmu.edu/>

U. Pitt. Health Sciences Library System <http://www.hsls.pitt.edu/>

Byzantine Catholic Seminary Library <http://www.bcs.edu/seminary/library.php>

\*We also used <http://www.compete.com> to help gather data.