

Assignment 3: UNESCO World Heritage Travel Database

LIBR 202-15

Sarah H. Scott

San José State University

## **UNESCO World Heritage Travel Database**

### **User Group**

The primary user group consists of travel agents working for the AAA Travel Agency, which specializes in travel in Europe and North America. The agents create custom travel itineraries and make travel arrangements for their clients and also collaborate in the creation of marketing materials and brochures for periodic travel promotions offered by the agency. All of the agents have at least a high school diploma, and many of them are college-educated. They have all been trained in the use of computer software, including a number of databases, necessary for the performance of their work. A secondary user of the database is the agency's receptionist, who assists with such tasks such as proofreading, copy-editing, and layout of the marketing materials produced by the agents; her primary purpose in using the database is for fact-checking and spelling verification.

### **Database Purpose**

The AAA Travel Agency is doing a promotion featuring UNESCO World Heritage Sites in Europe and North America. The database includes photos and information that travel agents can use, along with other resources available to them, in the creation of marketing materials and brochures for their promotion. It also allows agents to perform queries using specific criteria to find World Heritage Sites that will meet their clients' travel needs; for example, for a client who wants to take his family on a cultural tour of Europe, an agent can search for cultural sites in Europe, or for a client who has a lot of Delta Air Lines frequent flier miles, an agent can search for sites accessible directly from the U.S. via that airline. In addition, the database includes information useful for clients planning trips to particular sites, such as visa requirements, travel

warnings, and local climates, currencies, languages, and religions. Because the agents have a variety of resources at their disposal for use in creating their promotional materials, the level of detail offered by the database is sufficient. The information contained in the database comes either directly from UNESCO's World Heritage List Web site or from other authoritative sources such as the U.S. Department of State, so the agents can be confident that the information is reliable; and controlled vocabularies applied to most fields allow users to assume that all of the World Heritage Sites in the database that are related to a particular search term will be retrieved together.

**DATA STRUCTURE**

FIELD	NAME	DBTEXTRULE	ASSIGNMENT COMPLIANCE	INSTRUCTIONS FOR DATA ENTRY
1	<b>Record Number</b>	Rule 1. Automatic number (use default for automatic number definition) Rule 2. Term indexed Rule 3. N/A Rule 4. N/A Rule 5. N/A Rule 6. N/A	Control field	Automatically generated.
2	<b>Name of WH Site</b>	Rule 1. Text Rule 2. Term indexed and keyword indexed Rule 3. Required Rule 4. Unique entry only Rule 5. Single entry only Rule 6. No content validation	Identifier	Use name of World Heritage Site as found on <a href="http://whc.unesco.org/en/list">http://whc.unesco.org/en/list</a> .

3	<b>Date of Inscription</b>	<p>Rule 1. Date Rule 2. Term indexed and keyword indexed Rule 3. Required Rule 4. Not unique entry Rule 5. Single entry only Rule 6. No content validation</p>	Chronological	<p>Enter date only as 4-digit year as found on <a href="http://whc.unesco.org/en/list">http://whc.unesco.org/en/list</a>.</p> <p>Examples: 1979, 2001, 2008.</p>
4	<b>Type of Site</b>	<p>Rule 1. Text Rule 2. Term indexed Rule 3. Required Rule 4. Not unique entry Rule 5. Single entry only Rule 6. Use content validation</p>	Subject	<p>As indicated on <a href="http://whc.unesco.org/en/list">http://whc.unesco.org/en/list</a>, describe the World Heritage Site. Choose one from validation list: Cultural site, Natural site, Mixed site.</p>
5	<b>Country</b>	<p>Rule 1. Text Rule 2. Term indexed and keyword indexed Rule 3. Required Rule 4. Not unique entry Rule 5. Not single entry Rule 6. No content validation</p>	Geographic, controlled	<p>Use the name of the country in which the World Heritage Site is located as found on <a href="http://whc.unesco.org/en/list">http://whc.unesco.org/en/list</a>.</p>

6	<b>Continent</b>	<p>Rule 1. Text Rule 2. Term indexing Rule 3. Required Rule 4. Not unique entry Rule 5. Single entry only Rule 6. Use content validation</p>	Student's choice, Validation list	<p>Enter the continent in which the World Heritage Site is located. Choose one from validation list: Europe, North America.</p>
7	<b>Climate</b>	<p>Rule 1. Text Rule 2. Keyword indexed Rule 3. Not required Rule 4. Not unique entry Rule 5. Not single entry Rule 6. No content validation</p>	Student's choice	<p>Enter climate terms for the area in which the World Heritage Site is located as found on <a href="http://www.state.gov/r/pa/ei/bgn/">http://www.state.gov/r/pa/ei/bgn/</a>.</p> <p>Examples: Temperate, Mediterranean, Arctic.</p>
8	<b>Geographical Setting</b>	<p>Rule 1. Text Rule 2. Term indexing and keyword indexing Rule 3. Required Rule 4. Not unique entry Rule 5. Not single entry Rule 6. No content validation</p>	Geographic, uncontrolled	<p>Enter topographical terms describing the site and its surrounding area. Enter terms in singular form. Do not capitalize first letters of terms.</p> <p>Examples: forest, river, mountain.</p>

9	<b>Activities</b>	Rule 1. Text Rule 2. Term indexing and keyword indexing Rule 3. Not required Rule 4. Not unique entry Rule 5. Not single entry Rule 6. No content validation	Non-UNESCO data	Control using Library of Congress Subject Authorities ( <a href="http://authorities.loc.gov/cgi-bin/Pwebrecon.cgi?DB=local&amp;PAGE=First">http://authorities.loc.gov/cgi-bin/Pwebrecon.cgi?DB=local&amp;PAGE=First</a> ). List activities available at or within a one-mile radius of the World Heritage Site.  Examples: Tourism, Shopping, Worship, Dinners and dining.
10	<b>Languages Spoken</b>	Rule 1. Text Rule 2. Term indexing and keyword indexing Rule 3. Required Rule 4. Not unique entry Rule 5. Not single entry Rule 6. No content validation	Non-UNESCO data	Enter languages commonly spoken in the country in which the World Heritage Site is located as found on <a href="http://www.state.gov/r/pa/ei/bgn/">http://www.state.gov/r/pa/ei/bgn/</a> . Control using Library of Congress Subject Authorities ( <a href="http://authorities.loc.gov/cgi-bin/Pwebrecon.cgi?DB=local&amp;PAGE=First">http://authorities.loc.gov/cgi-bin/Pwebrecon.cgi?DB=local&amp;PAGE=First</a> ).

11	<b>Local Religions</b>	Rule 1. Text Rule 2. Term indexing and keyword indexing Rule 3. Not required Rule 4. Not unique entry Rule 5. Not single entry Rule 6. No content validation	Non-UNESCO data	Enter religions commonly practiced in the country in which the World Heritage Site is located as found on <a href="http://www.state.gov/r/pa/ei/bgn/">http://www.state.gov/r/pa/ei/bgn/</a> . Control using Getty Art & Architecture Thesaurus ( <a href="http://www.getty.edu/research/conducting_research/vocabularies/aat/">http://www.getty.edu/research/conducting_research/vocabularies/aat/</a> ).
12	<b>Local Currency</b>	Rule 1. Text Rule 2. Term indexing and keyword indexing Rule 3. Required Rule 4. Not unique entry Rule 5. Not single entry Rule 6. No content validation	Non-UNESCO data	Enter the currency of the country in which the World Heritage Site is located. Control using ISO 4217 currency names ( <a href="http://www.iso.org/iso/support/faqs/faqs_widely_used_standards/widely_used_standards_other/currency_codes/currency_codes_list-1.htm">http://www.iso.org/iso/support/faqs/faqs_widely_used_standards/widely_used_standards_other/currency_codes/currency_codes_list-1.htm</a> ).

13	<b>Airlines</b>	<p>Rule 1. Text Rule 2. Term indexing and keyword indexing Rule 3. Not required Rule 4. Not unique entry Rule 5. Not single entry Rule 6. No content validation</p>	Non-UNESCO data	<p>List airlines that fly directly from the United States to the country in which the World Heritage Site is located as found on <a href="http://www.skyscanner.com/">http://www.skyscanner.com/</a>. If the site is located in the United States, leave blank. If there is no airline with a direct flight from the United States, enter "No direct flight." Control using Library of Congress Subject Authorities (<a href="http://authorities.loc.gov/cgi-bin/Pwebrecon.cgi?DB=local&amp;PAGE=First">http://authorities.loc.gov/cgi-bin/Pwebrecon.cgi?DB=local&amp;PAGE=First</a>).</p>
14	<b>Visa Required</b>	<p>Rule 1. Text Rule 2. No indexing Rule 3. Not required Rule 4. Not unique entry Rule 5. Single entry only Rule 6. Use content validation</p>	Validation list	<p>If a visa is required to enter the country in which the World Heritage Site is located, enter Yes. If not, enter No. Choose one from validation list: Yes, No.</p>
15	<b>Travel Warning</b>	<p>Rule 1. Text Rule 2. No indexing Rule 3. Required Rule 4. Not unique entry Rule 5. Single entry only Rule 6. Use content validation</p>	Non-UNESCO data, Validation list	<p>If the U.S. Department of State (<a href="http://travel.state.gov/travel/cis_pa_tw/tw/tw_1764.html">http://travel.state.gov/travel/cis_pa_tw/tw/tw_1764.html</a>) has issued a Travel Warning for the country in which the World Heritage Site is located, enter Yes. If not, enter No. Choose one from validation list: Yes, No.</p>

16	<b>Photo of Site</b>	Rule 1. Image Rule 2. No indexing Rule 3. N/A Rule 4. N/A Rule 5. N/A Rule 6. N/A	Student's choice	Use only one JPEG image from Gallery tab in World Heritage Site record on <a href="http://whc.unesco.org/en/list/">http://whc.unesco.org/en/list/</a> .
----	----------------------	--	------------------	---

## Log

As our class embarked on Assignment 3, several of my teammates and I decided to work collaboratively on developing a database purpose, user group, and data structure and to share the work of data collection. Our work group consisted of A.B., J.L., J.W., and me. J.L. proposed our chosen user group and database purpose, and once we had agreed upon it, we used our team discussion forum to develop a work plan. Our first task was to define our database fields. Since some of the fields had to fulfill specific requirements, defining many of them was fairly straightforward. Others required more thought and creativity. We decided to include information not only about the World Heritage Sites themselves but also travel information that would be useful in planning trips; for example, we included a Local Religions field because we recognized that travelers might need to pack special attire when visiting locations in which certain religions, such as Islam, are commonly practiced. The field that was the most challenging for us to define was the one that would fulfill the requirement for an uncontrolled geographic field. Many of our initial ideas, such as climate or place names, seemed to work better as controlled fields. We finally decided to call the field “Geographical Setting” and populate it with topographical terms describing the World Heritage Sites and their surrounding areas. This information is useful for our user group because they might have clients interested in visiting areas with specific features such as mountains or islands, but the use of a validation list or an external controlled vocabulary is not necessary.

Once we had agreed upon most of the fields for our database, our group narrowed down the multitude of World Heritage Sites to 50 by using the Advanced Search function on the UNESCO Web Site to search for “Sites with photo gallery” in “Europe and North America.” This brought up a list of exactly 50 countries, many of them with multiple World Heritage Sites

listed under them. We initially chose one site in each country, striving to select a mix of natural, cultural, and mixed sites. J.L. created an Excel spreadsheet listing the 50 sites along with columns for all of the other fields we had selected for our database. We somewhat arbitrarily divided the work of data collection amongst the four of us, with each of us committing to collect data for 12-13 sites. In addition, one of our teammates, A.B., volunteered to collect information on visa requirements for all of the sites.

Before we began to collect data, we realized that it was important to create at least a draft of our data structure to guide our data collection efforts, allowing us to enter the data directly into the spreadsheet according to pre-selected rules. We also recognized the value of getting feedback from our classmates and instructor to reassure us that we were on the right track. Taking into account the requirements of the assignment, our team discussion, and the discussion that had already taken place on our class section's Assignment 3 Discussion forum, I filled out the data structure template. For each field, I thought carefully about how to define the six required rules. For some fields, such as the automatically generated record number, the nature of the data made it easy to define the rules; however, most of the fields required more thought. For those fields that fulfilled specific requirements of the assignment, once we had chosen the field name and the nature of the data based upon the needs of our user group, it was only a matter of deciding how to apply the rules. Indexing decisions were made based upon how the user group would likely be querying the database. For example, I decided to term index the Continent field because I thought users might want to search for all of the sites in Europe or all of those in North America; I keyword indexed the Climate field because I thought users would be likely to search using terms such as "continental" or "temperate" but not necessarily "continental temperate." For those fields with a small number of possible entries, such as "yes" and "no," I decided to use

validation lists. The decision of whether to make a field required or repeatable was usually dictated by the data. For example, if there was no question that data was available, I made the field required, whereas it would otherwise not be required. If there was only one possible entry that would fit, I made it non-repeatable, whereas I made it repeatable if I knew there would be multiple possible entries. Most fields were defined as “not unique entry” with the exception of the Name of WH Site field, since each site could be listed in only one record.

Because of the importance of consistency in the database, we chose to apply controlled vocabularies to many of our database fields. The particular controlled vocabulary applied to a specific field was chosen based upon the nature of the data. For example, after learning from our instructor about the ISO 4217 currency names, we chose that list as the controlled vocabulary for our Local Currency field. In the case of our Country field, which fulfilled the controlled geographic field requirement, it made sense to use the UNESCO Web site itself as the controlled vocabulary. However, in some cases, it is entirely possible that another controlled vocabulary would have worked equally well or possibly even better. For example, early on, we decided to use the Getty Art & Architecture Thesaurus to control our Local Religions field. Since religious terms are relevant to the fields of art and architecture, we were able to successfully utilize the Getty thesaurus for our purposes. However, the Library of Congress Subject Authorities probably would have worked just as well. We also encountered some challenges with controlled vocabularies. As I posted in the Assignment 3 Discussion forum, I couldn't always find a Library of Congress subject heading for every term I needed, and sometimes the subject headings I did find seemed awkward, as when I wanted to list “skiing” as an activity available at or near some of the World Heritage Sites, but was forced to choose “Skis and skiing” from the list of approved subject headings. Our instructor offered some helpful suggestions for these

problems. One option she suggested for the problem of not being able to find an appropriate Library of Congress subject heading was to use an alternate controlled vocabulary more appropriate to the data. For the problem of awkward headings, she acknowledged that this is a major challenge faced by catalogers on an ongoing basis and offered several suggestions, such as using the best available heading or creating a validation list. In the end, I decided to use the best available heading. I thought it was important to use consistent terms, and while the Library of Congress headings were sometimes awkward, at least they did encapsulate the ideas I needed to convey through the data.

Once I had completed the first draft of our data structure, I posted it on the Assignment 3 Discussion forum for feedback. At that point, my group still hadn't decided on a field to fulfill the uncontrolled geographic requirement, and the feedback we received from our instructor was very helpful. After some additional discussion, we made the final decision on that field, set our team data collection deadlines, and set out to collect our data.

As I began to collect my share of the data, I quickly realized that it would be to our advantage to identify some data sources besides the UNESCO Web site from which we could gather data that would be authoritative and consistent. We found the travel section of the U.S. Department of State Web site ([www.state.gov/travel](http://www.state.gov/travel)) to be a good source for information on visa requirements and travel warnings, and its Background Notes section ([www.state.gov/r/pa/ei/bgn/](http://www.state.gov/r/pa/ei/bgn/)) provided us with a wealth of information on climate, religions, and languages of the countries in which the World Heritage Sites are located. Wikipedia ([www.wikipedia.org](http://www.wikipedia.org)) provided us with information about activities and geographical features, and Skyscanner ([www.skyscanner.com](http://www.skyscanner.com)) was a great source for finding airlines with direct flights from the U.S. to the countries in which the World Heritage Sites are located. As I entered my

data into the spreadsheet, I kept the data structure close at hand and made sure to follow the rules and instructions for data entry. For the most part, my data collection process went fairly smoothly. In a handful of instances, when collecting data for a controlled field, I was unable to find an authorized term or subject heading. For example, I was unable to find a listing for the Druze religion, practiced in Israel, in the Getty Art & Architecture thesaurus. Since it is a minority religion, I decided not to enter it at all rather than have to scrap all of my previous work and choose a whole new controlled vocabulary. However, I recognize that this is one of the major challenges of working with controlled vocabularies. By omitting information, as the person entering the data into the database, I did a disservice to the user group. I can't claim to know the solution to this problem, but I do hope that the keepers of controlled vocabularies such as the Library of Congress and the Getty Vocabulary Program allow for the admission of new terms as needed.

Once everyone on my team had completed his or her data collection responsibilities, we posted our data on our team discussion forum, and J.L. compiled all of the data in a single spreadsheet. Not everyone had entered all of their data according to the rules and instructions for data entry specified in our data structure, so we spent a couple of days revising the data. In the end, I also had to "clean" some of the data gathered by my teammates as I entered it into my database, because there were certain inconsistencies in capitalization and grammar as well as varying levels of specificity in the data. This experience and my observations working with shared, open databases in previous jobs suggest to me that attention to detail and consistency in the application of data entry rules are vitally important to the integrity of a database. The higher the number of individuals entering data, the higher the risk of inconsistencies. For this

assignment, as the database designer, it was relatively easy to clean the data and eliminate inconsistencies. However, in a large, open database, the challenge would be much greater.

Once our group's data was compiled, our work as a group was complete and we were ready to create our databases in DB/TextWorks. However, before I began, I made some additional changes to my data structure. Our instructor had suggested in her comments on our draft data structure that we reorder the fields and group related fields together, so I did so, moving from information about the World Heritage Sites and their geography to travel-related information and ending with the photo field. I also went over the rules and instructions for data entry one last time. Once I was satisfied with the data structure, the process of creating my database in DB/TextWorks began. Because I had done so much of the work of creating fields and rules beforehand, creating the database itself was fairly easy. I had watched the Elluminate tutorial on using DB/TextWorks and referred closely to our lecture notes from Week 9 on working with DB/TextWorks. One challenge that I did encounter, which had been discussed by other classmates in the Assignment 3 Discussion forum, was a limit on the number of characters that could be used for a field name. With a little creative rephrasing and abbreviation, I was able to enter suitable field names. When I had entered all of the fields and their rules into DB/TextWorks and the database structure was complete, I started to enter the records. Because of the need to "clean" some of the data, as discussed above, I decided not to import the data directly from Excel but rather to enter it manually. Although my manual process was more time-consuming than importing the data, I thought it was necessary to ensure the consistency and integrity of the data. After entering the first few records, I ran a few queries on the database, searching, for example, for all the records with Europe listed in the Continent field and all the

records with Roman Catholicism listed in the Local Religions field. Satisfied that my queries were retrieving the desired data, I completed the entry of 30 records into the database.

Once all of the records had been entered into the database, I tested it by devising several more queries. First, I searched for all of the cultural sites in Europe (Figure 1), which retrieved 16 records. Figure 2 shows, in Report View, the first of the 16 records along with the image associated with the file path in the Photo of Site field for that record. Another query I ran was for all records with “skiing” in the Activities field. This query retrieved 2 records, the first of which is shown, along with its associated image, in Figure 3. It is significant to note that although the actual subject heading listed is “Skis and skiing,” because the Activities field was keyword indexed, I was able to successfully retrieve all of the sites in the database where skiing is listed by searching with just the word “skiing.” I did have a little difficulty in devising a

**Figure 1: Query for Cultural sites in Europe**

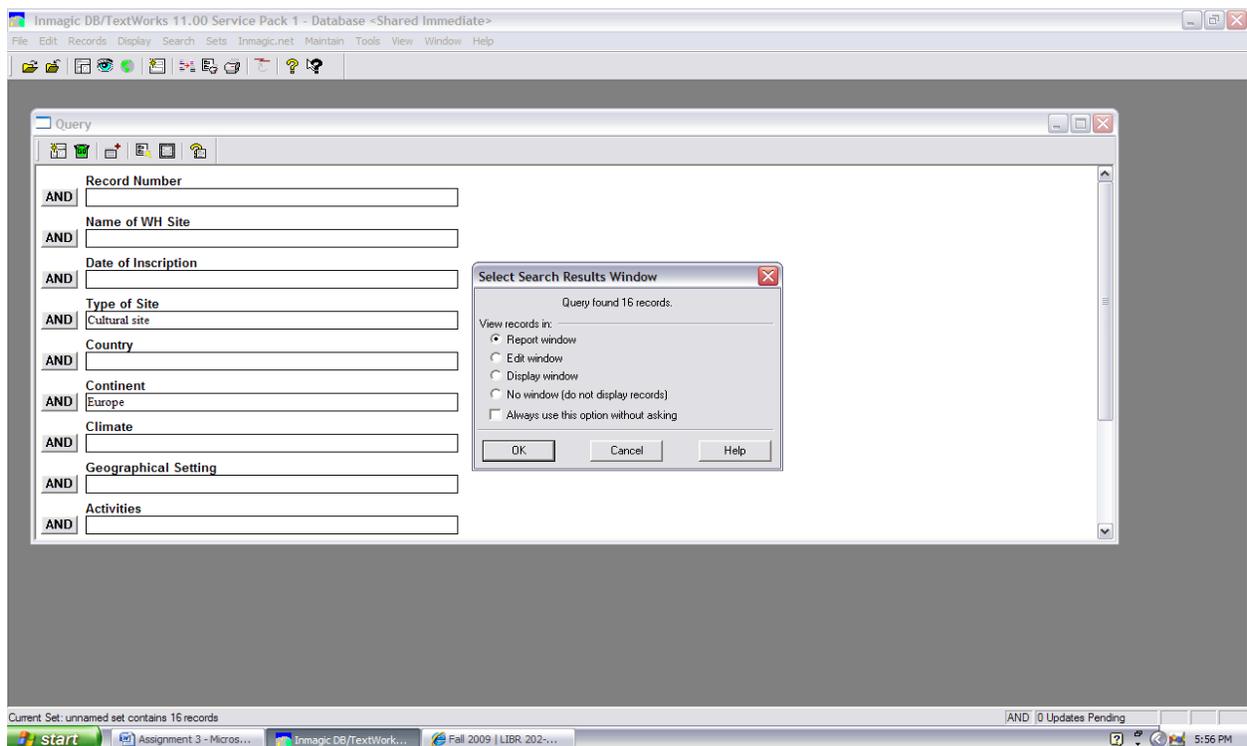


Figure 2: Record 1 of 16 retrieved by query for Cultural sites in Europe

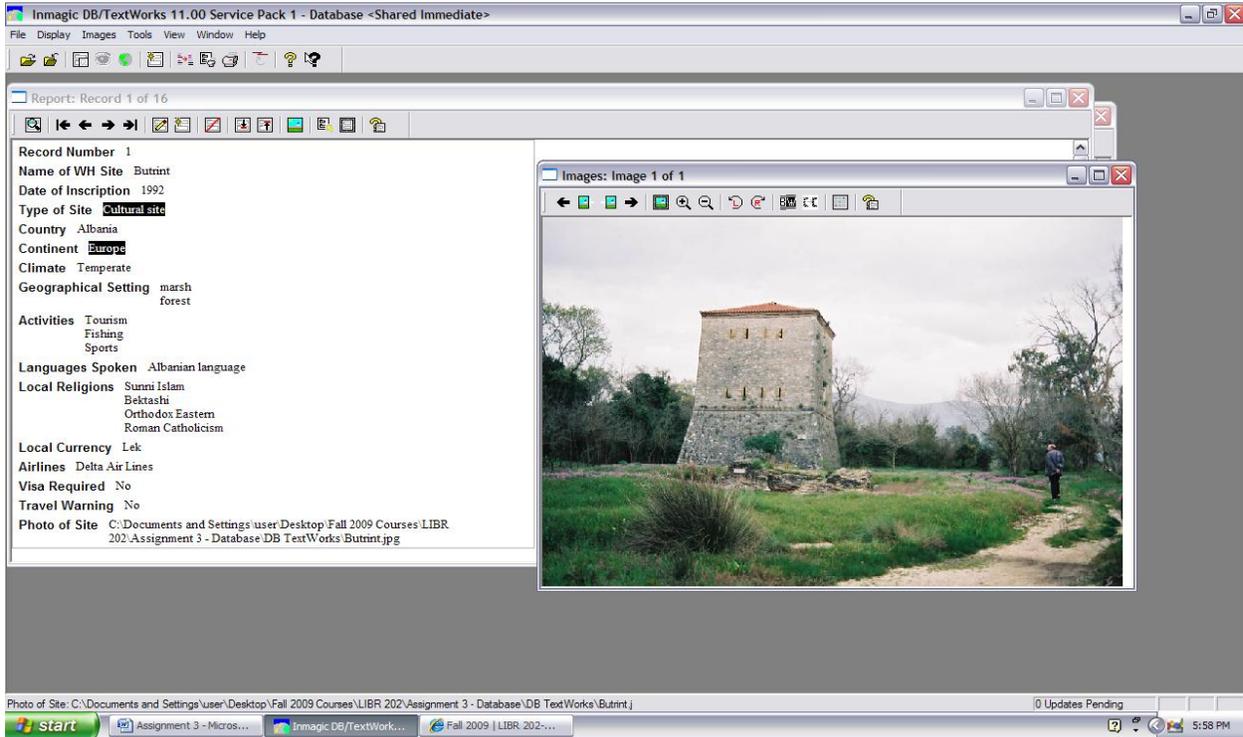
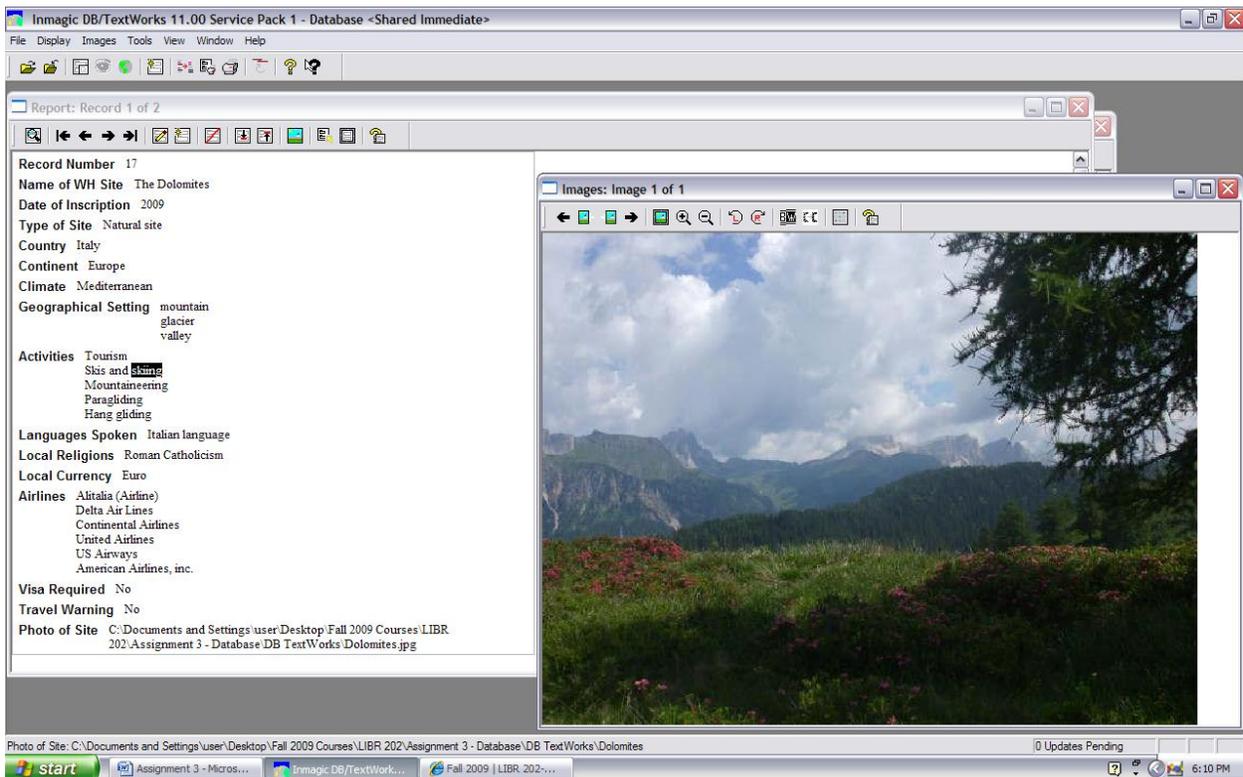


Figure 3: Record 1 of 2 retrieved by query for “skiing” in Activities field



successful query by Date of Inscription. I had chosen to make the field both term and keyword indexed, but I initially had trouble retrieving any records unless I searched by a specific year. However, I finally used the Help function in DB/TextWorks to search for “Truncation Searches,” and I learned that by using the search term “19\*” I could retrieve all of the records for sites inscribed in the 1900s, and by using the search term “200\*” I could retrieve all of the records for sites inscribed between 2000 and 2009.

### **Evaluation**

Overall, I think my database turned out very well. I think that my user group will find it to be both easy to use and an authoritative and accurate source of information for their work. Because it was designed with their information needs, backgrounds, and usage styles in mind, they should have little trouble in retrieving the information they seek. While I did encounter some challenges in working with the controlled vocabularies, such as awkward subject headings, not always knowing which controlled vocabulary to use, and a small number of unrepresented terms, overall, I think that the use of controlled vocabularies helps to ensure consistency among all of the records. I think that my group’s choice of fields is very appropriate to the needs of our user group and to the requirements of the assignment. I also think that my choices in the application of rules to the various fields are appropriate and effective, as my various queries confirmed. I learned a lot from this assignment and enjoyed the challenge of creating a database from scratch, and as a result, I have a much greater understanding of the many challenges that go into creating a successful database.