

Worcester Polytechnic Institute Digital WPI

Interactive Qualifying Projects (All Years)

Interactive Qualifying Projects

March 2009

Cataloging 19th Century Periodical Images

Candace Rachel Chouinard
Worcester Polytechnic Institute

John Forrest Hogeboom
Worcester Polytechnic Institute

Sarah B. Fredette
Worcester Polytechnic Institute

Steven M. Barton
Worcester Polytechnic Institute

Follow this and additional works at: <https://digitalcommons.wpi.edu/iqp-all>

Repository Citation

Chouinard, C. R., Hogeboom, J. F., Fredette, S. B., & Barton, S. M. (2009). *Cataloging 19th Century Periodical Images*. Retrieved from <https://digitalcommons.wpi.edu/iqp-all/2506>

This Unrestricted is brought to you for free and open access by the Interactive Qualifying Projects at Digital WPI. It has been accepted for inclusion in Interactive Qualifying Projects (All Years) by an authorized administrator of Digital WPI. For more information, please contact digitalwpi@wpi.edu.

WORCESTER POLYTECHNIC INSTITUTE

Cataloging 19th Century Periodical Images

A Term 2008 – C Term 2009

Steven Barton

Candace Chouinard

Forrest Hogeboom

Sarah Latta

3/13/2009

Abstract

This project created a database of images in nineteenth-century *Scientific American* issues, including relevant information as well as categorizing by subject matter. Sponsored by the American Antiquarian Society, the project developed an online electronic database using MySQL. In addition, the entries created by the previous team were edited, and volumes 10-12 were added to the database.

Table of Contents

1. Introduction	5
2. Our Sponsor	6
3. Purpose of Project.....	6
4. Background.....	6
4.1. Previous Research	7
4.2. History of Patents and Scientific American	7
5. Cataloging.....	11
Previous Headings	11
New Headings	12
Example	17
6. Database.....	18
1.1. Initial Implementation	19
Final Implementation	20
How it works.....	24
7. Future Work.....	26
8. Conclusions	27
9. Citations.....	28

Authorship

Introduction	Sarah Latta
Our Sponsor	Candace Chouinard
Purpose of Project	Sarah Latta
Background	
Previous Research	Sarah Latta
Patent History	Candace Chouinard
Cataloging	Steven Barton
Database	John Forrest Hogeboom
Future Work	Sarah Latta
Conclusions	Candace Chouinard

Editing: Candace Chouinard and Sarah Latta

1. Introduction

The American Antiquarian Society saw a need for an electronic database of 19th century engravings from periodicals. This idea was brought to Professors Bullock and Samson who created an IQP to address the issue in 2007. The group that designed and established the database from the fall of 2007 to spring 2008 accomplished much. They determined what periodical to use, what cataloging format, and catalogued eight volumes. Images from *Scientific American* published during 1844 and 1854 were catalogued using a modified Dublin Core method.

When our group started the project we met several obstacles. The database had previously been hosted on a student's website that was no longer available. Large amounts of time were spent just trying to access the previous database. When access was available to the raw computer code, it was hard to read and nearly impossible to follow. One group member took charge of all database related issues and crafted a new database that could be read, accessed, and edited by all.

The cataloging process used by the previous group was confusing to follow and, without being able to view the catalog, nearly impossible to understand. One of the previous group's members met with us and answered many of our questions. Based on that meeting, we modified the cataloging headings, making them easier to understand. After establishing this format, we moved on to cataloging and completed volumes 10 - 12.

2. Our Sponsor

The American Antiquarian Society is located right down the street from Worcester Polytechnic Institute in Worcester, Massachusetts. The research library was founded in 1812 and remains a vital resource of historical research materials. The AAS collection spans from the colonial era to reconstruction, and includes manuscripts, pamphlets, books, and periodicals like *Scientific American*. A center for researchers, the American Antiquarian Society continues to preserve physical artifacts while modernizing its resources using digital catalogs.

3. Purpose of Project

The purpose of this project is to continue developing a searchable database of nineteenth-century images. To date we have focused on *Scientific American*, though in the future this could be expanded. Cornell University's "Making of America" provides an online version of the magazine that allows users to search text, but not images. Our database entries contain relevant keywords, inventor information, and an article summary--all important parts of an effective image search library. For instance, if a historian is writing a history of windmills, he or she would be able to search *Scientific American* for articles and images relating to windmills.

4. Background

The team from last year provided a solid foundation in understanding the history of *Scientific American* and other relevant topics. Below is a relevant summary of their research, as well as patent history and how *Scientific American* shaped it.

4.1.Previous Research

Scientific publications began in 1665, around the beginning of the Enlightenment. Originally, there were science articles included in scholarly journals, but as scientific pursuits increased, the size and number of science-related articles increased. Until *Scientific American*, science and scholarly periodicals were geared toward an upper class considered to have the time to pursue knowledge. *Scientific American* was the first science related periodical to be accessible and designed for individuals whom though interested, may not have had a formal education. Early periodicals were frequently publishes as serialized books. Binding books, and high quality printing, possibly including color illustrations, cost quite a bit. *Scientific American* was published as a newspaper that required little binding and used wood cut engravings to keep the cost down. Also, *Scientific American* included sections geared to teach different ideas and concepts rather than just reporting new inventions.

4.2.History of Patents and Scientific American

Rufus Porter, an American inventor, presented his first publication of "The Advocate of Industry and Enterprise, and Journal of Mechanical and Other Improvements" in 1845. Porter instigated one of the most influential periodicals in the history of American inventing. His paper, later published under new management as *Scientific American* magazine, facilitated innovation through communication in the early scientific community. The periodical was integral in shaping the United States patent system; not only did *Scientific American* owners Munn & Co financially and intellectually support rising inventors, but the company also founded the most successful patent review agency in history ("About Us"). *Scientific American* contributed significantly to the growth of American industry through facilitating, reviewing, and advertising patented innovations.

Before the influence of *Scientific American*, the young American government instituted fundamental patent legislation to mimic European precedents. English Parliament passed the Statute of Monopolies in 1624, enabling limited grants of exclusivity to some inventions. The statute specified that inventions must be novel (“A History”). Some state constitutions within the U.S. provided for patents under similar conditions, the first given in Massachusetts to Samuel Winslow in 1641 for a method of salt production (Bellis). The United States Constitution, ratified in 1790, established national protection of patented products. Article 1, Section 8, clause 8 states, “Congress shall have power ... to promote the progress of science and useful arts by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries” (“A History”). For the first time in U.S. History, innovators had legal rights to their product. The Constitution provided the groundwork for more specific regulation by guaranteeing the authority of Congress to grant patents, a powerful catalyst for the industrial revolution. The American government drew from a long history of European patent regulation to establish a fundamental right to intellectual property.

Though the United States Constitution was a critical step toward national regulation, numerous Patent Acts followed to match legislative growth to the country’s economic success. The U.S. issued three patents in 1790 compared to nearly 500 in 1854 (“US Patent”). Patent Office practices changed frequently during this period, at times for the worse. The Patent Law of 1790 gave only the Secretary of State, Secretary of War, and Attorney General the authority to grant patents. The inventor needed to prove the novelty, usefulness, and reproducibility of his innovation to a patent officer and could earn up to 14 years of exclusive production rights. Thomas Jefferson redefined patentable inventions in the Patent of Act of 1793, “...any new and useful art, machine, manufacture or composition of matter and any new and useful improvement

on any art, machine, manufacture or composition of matter” (“A History”). The 1793 Act legislated additional protection for inventions later improved upon. The inventor of the elaborated product had no rights to its predecessor. Indicative of the social climate, the 1793 Patent Act also limited grants to United States citizens. More significantly, this new act restructured the patent review system to alleviate demands on high officials (Jones 7). Patents were not “examined”; instead, an emerging Patent Office loosely processed registrations. Dr. William Thornton, appointed chief officer from 1802 to 1828, was a brilliant inventor but proved an irresponsible official. Thornton made arbitrary decisions about fees and often neglected record keeping; discrimination and favoritism were rampant (Jones 8). Thornton’s successor, Dr. John D. Craig, furthered corruption of the office with his “tyrannical and unjust” (Jones 11) practices. Further legislative adjustments were clearly needed to restore the integrity of the patent system.

The Patent Act of 1836 reformed the application process by re-instating a requirement for thorough examinations. The Patent Office evolved accordingly to include a chief clerk, examiner of patents, a machinist, three additional clerks, and a messenger (Jones 14). The number of patent applications quickly overwhelmed the small offices and independent examiners stepped in. Patent lawyers would authorize an application if it followed conditions of the law. Federal regulators relied on judgments of these lawyers when granting patent rights (Dobyns). Predictably, this compromised the fairness of the grant process as the Patent Office did not have sufficient oversight of third party reviewers. Progressing legislation, even during the late nineteenth century, was often not specific enough to ensure fair review of patent applications. Between 1790 and 1860, 765 court cases involved contesting patent rulings. This number does not seem significant until one considers that the U.S. Court system handled an average of only

380 cases per year in this same period (Khan “Property”). Inconsistency among overburdened reviewers was a recurrent problem even as lawyers and Patent Office officials were employed under stricter legislation.

Alfred Ely Beach and Orson Desaix Munn, the same businessmen who purchased *Scientific American* in 1846, saw entrepreneurial potential in the strained patent system. They founded Munn & Co shortly after acquiring *Scientific American* and diversified to establish a legal office. The reign of Beach and Munn’s office began when Allen Benjamin Wilson came to Munn & Co. in New York City. Seeking help with obtaining patents, he brought models of a sewing machine and rotary steam engine. Munn & Co. patented his sewing machine the same year and Mr. Wilson grew wealthy from selling his ingenious design (Dobyns). The Munn & Co firm quickly dominated the patent review industry, handling one-third of all patents issued in the years around 1860. Even though the office closed in 1960, Munn & Co was responsible for issuing 3.5 percent of all patents to date (Dobyns). Munn & Co.’s patent office was an invaluable resource for aspiring inventors lacking the knowledge or funds to capitalize.

Serving to advertise products patented at Munn & Co. and inspire other developments, the now widely circulated *Scientific American* spread knowledge of the latest technology to layman across the country. Munn and Beach’s paper was devoted to covering the latest patents and inventions. As the oldest continually published periodical in the country, it was the first journal to effectively publicize patent claims (Dobyns). The demand for page space was so high by 1876 that Munn & Co. published a supplementary booklet (“Orson”). *Scientific American* played more than the passive role of modern media; the early journals were an important catalyst in the Industrial Revolution because of associated legal support and publicity for innovators.

5. Cataloging

Previous Headings

The below set of headings was devised by the 2007-2008 group and it is related to the Dublin Core method of cataloging. The main differences between the two sets regard certain additions the group made, including the second description element concerning machine readable inventions and the comment element. Thus one can think of this metadata element set as a more detailed version of the Dublin Core metadata element set:

Title:

Description:

Keywords:

Publication: sciam

Volume:

Date:

Issue:

Page:

Column:

Subject:

Description:

Creator:

Genre:

Person:

Task:

Series:

Description:

Comment:

The title heading contains the title of the image being cataloged and if there is not a title then this field contains the title of the article. The first description element contains a description of the image. This usually entails what type of view the image is, such as a plan view, or a perspective view. Keywords are either of the illustration or the subject, without distinction. The information necessary to track the image is cataloged under the publication, volume, date, issue, page, and column fields. When cataloging the column, the leftmost column is used. The devices or inventions and their creators are cataloged under the subject, description, genre, and creator field respectively. The subject field contains the devices name. The second description heading contains what the subject of the device is. However if the subject of the device is self-explanatory the subject description is omitted. The creator's, or the inventor's name is kept track of and the genre the subject of the device is about is also kept track of with no standardized genre titles. The person who created the illustration and the task they completed are cataloged under the person and task elements. The person element contains the artists name and the task element contains what the person did on the illustration. If the images appear in a series then they are cataloged under the series and third description heading. The series name goes under the series heading and a description of the series is cataloged under the third description heading. The description of the series usually contains what the purpose of the series is.

New Headings

Upon examining the above headings, our group noticed that some of the elements were convoluted, redundant, and unnecessary for the success of our database's search function. After

discussing the issue with Allison Smyth of last year's group, our team decided to research other options. The Library of Congress formatting, known as MARC, was investigated at the request of our sponsor, the American Antiquarian Society. MARC is machine readable and thus would facilitate integration of our database into a larger system. We met with Lora Brueck and Rodney Obien, two librarians at Worcester Polytechnic Institute's Gordon Library, to discuss the plausibility of using MARC. Though we all agreed that MARC would be ideal with respect to compatibility, the logistics of implementing the system are overwhelming. MARC requires assigning several four digit numbers to different levels of the entry to uniquely identify a catalog entry. These standards for these numbers and the overall cataloging process are very regulated and precise. The time it would take to learn this process to a reasonable level of accuracy would detract from work actually done. Given this, it was decided that utilizing MARC standards would not be feasible with our skill level and time constraints. The team finally agreed on a format similar to the Dublin Core format that the previous team used.

We removed some of the redundant and vague elements, such as column, subject, the second description, task, and comment from the previous format. The headings were simplified and made coherent. The new format can be seen below, followed by a description of each heading.

Title:

Image Description:

Artist:

Article Summary:

Inventor:

Author:

Patented:

Subject:

Publication:

Date:

Volume:

Issue:

Page:

Series:

Keywords:

Title: the procedure for obtaining this property is identical to last year's in that the illustration's title was used and if there was not a title of the illustration, the article's title was used.

Image description: identical to the first description element used by the previous group.

Therefore this element provides a description of the image/images, i.e. it is a perspective view, plan view, sectional view, etc

Artist: contains the artist's and/or the engraver's name. If both the artist and the engraver are known, they can be distinguished from one another using commas. For example, John Smith, engraver; Jack Hollis, artist.

Author: contains the author's name and where they were from, if listed. The author's name was inputted as it appeared in the article. If the article listed both first name and last name, it was entered that way—John Smith, of Norfolk, VA. If only initials were given, these replaced the first name.

Article summary: contains a summary of the article associated with the image. This would include any improvements that were made, and occasionally how it worked.

Inventor: contains the inventor's name and where they were from, if listed.

Patent: contains the patent date, if given, and was added to allow people to find more information on the inventions the images are showing.

Subject: provides information regarding the genre of the invention. In order to make the database search function work best we agreed on the following subjects obtained from a *Scientific American* article summarizing patents granted in 1854:

- | | |
|---|---|
| <input type="checkbox"/> Agricultural Implements and Operations | <input type="checkbox"/> Lever, Screw, and other Mechanical Power |
| <input type="checkbox"/> Metallurgy and Manufacture of Metals | <input type="checkbox"/> Grinding Mills and Mill Gearing |
| <input type="checkbox"/> Manufacture of Fibrous and Textile Substances | <input type="checkbox"/> Lumber, Machines for Manufacturing |
| <input type="checkbox"/> Chemical Process, Manufacture, and Composition | <input type="checkbox"/> Stone and Clay Manufactures |
| <input type="checkbox"/> Calorifics, Lamps, Stoves, etc | <input type="checkbox"/> Leather -- Tanning, Dressing and Manufacturing |
| <input type="checkbox"/> Steam and Gas Engines | <input type="checkbox"/> Household Furniture |
| <input type="checkbox"/> Navigation and Maritime Implements | <input type="checkbox"/> Arts -- Fine, Polite, and Ornamental |
| <input type="checkbox"/> Mathematical, Philosophical, and Optical Instruments | <input type="checkbox"/> Fire arms and Implements of War |
| <input type="checkbox"/> Civil Engineering and Architecture | <input type="checkbox"/> Surgical and Medical Instruments |
| <input type="checkbox"/> Land Conveyance | <input type="checkbox"/> Wearing Apparel and Implements for Manufacturing |
| <input type="checkbox"/> Hydraulics and Pneumatics | <input type="checkbox"/> Miscellaneous |

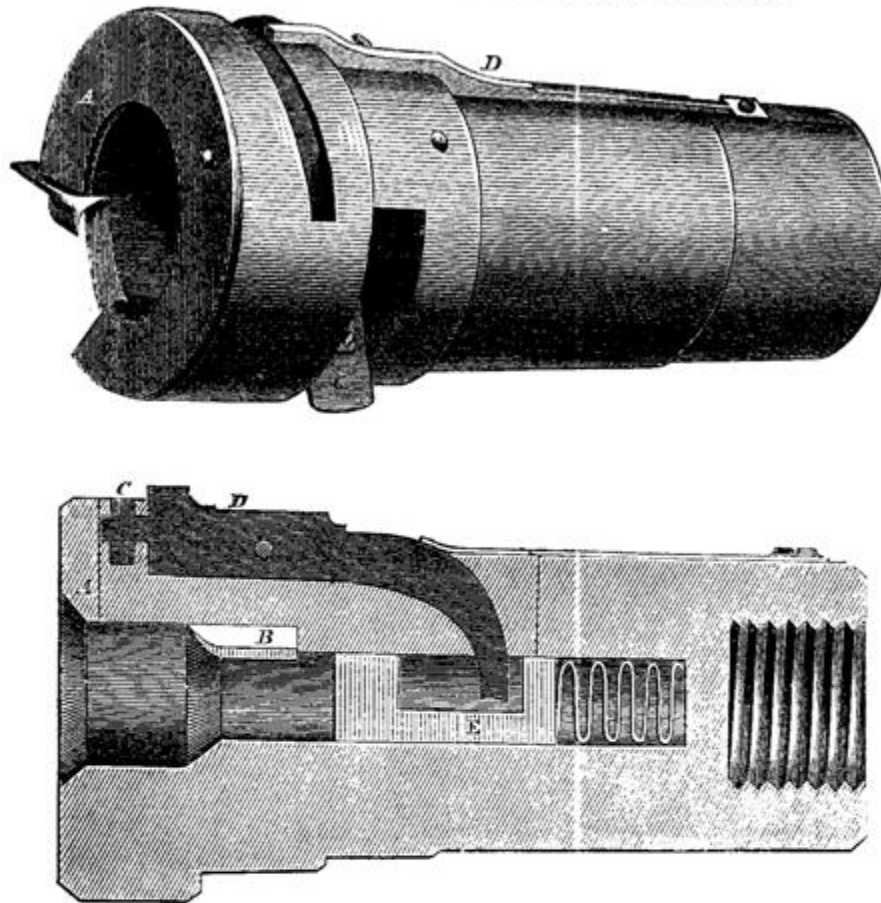
Series: a column or article that was part of a series was kept track of to arrange related images together and thus streamline the search function of our database. The series element would contain the title of the series.

Keywords: contains all of the data to allow searching on and thus is one of the most important elements. The data that comprised our keywords element came from the image itself or the article the image was based on.

Publisher, volume, date, issue, and page contain what they say. Publisher was always listed as Scientific American. Volume and issue numbers, and page number were included. Date gave the date that issue was published.

Example

CHUCK FOR MAKING BEDSTEAD PINS.



The accompanying figures are views of the improved Chuck of H. Gross, of Tiffin, Ohio for turning bedstead pins and tenons upon chair stuff.

Figure 1 is a perspective view, and figure 2 is a longitudinal vertical section of the chuck.

The body of the chuck is composed of cast-iron, in which three chisels are set. The first one at A, cuts the stick to the size required for the head of the pin; the second at B turns the shaft of the pin to the proper size, and the third chisel at C, attached to the end of lever, D, by pressure of the stick inward, is forced down, and turns the head of the pin, all of which is done without removal or change of the stick until the pin is completed. The end of the lever, D, opposite the chisel is made with a curved shank passing down into the orifice of the chuck, and is operated upon by a follower, E, connected with a spiral spring and against which follower the stick is press-

ed when being cut, forcing the lever upward at this end, while the chisel at the opposite end is brought down upon the stick at the proper time and place to form the head of the pin.

The chuck is so constructed as to be attached to any common lathe, or used separately, as may be desired, and with which one thousand pins per hour can be made by one person. By removing the lever chisel C, the chuck can be profitably used in turning tenons upon chair stuff; the two remaining chisels being set to turn the two sizes generally used, and which may be changed to larger or smaller size by the screw and slot in the chisels.

The patent for this improvement in chucks for turning bed pins was granted on the 23d of September last: the claim can be found on page 26, this volume, SCIENTIFIC AMERICAN. For further information apply to the patentee, Henry Gross, Tiffin, Ohio.

Title: Chuck for Making Bedstead Pins

Image Description: A perspective and longitudinal vertical view of the chuck.

Artist:

Article Summary: A versatile chuck that can attach to any common lathe.

Inventor: Henry Gross, Tiffin, Ohio

Author:

Patented: September 23, 1855

Subject: Lever, Screw and other Mechanical Power

Publication: Scientific American

Date: December 27, 1856

Volume: 12

Issue: 16

Page: 128

Series:

Keywords: bedstead, pin, chuck

As illustrated above, the cataloguing parameters are fairly intuitive and are well suited to characterize scientific images. A careful reading of the article and thorough examination of each image provides the cataloguer with the necessary information. Typed in a text file of the above format, the image data can then be uploaded to our database.

6. Database

The group last year successfully created a searchable database. However, when we began the project, we were not able to access, update, or even view it. Thus, the creation of a new database was necessary. Initially, the Python language based web framework Django was used. While it offered simplicity among several other advantages, it did not turn out to be feasible; the site was

remade with a series of other systems. Despite the significant revisions, there were always elements that remained the same, and the database in its current form represents a gradual progression from the initial concept.

1.1.Initial Implementation

Django framework is composed of a few components intended to be easily interchangeable. Models describe a database; the file is written in Python and contains objects with variables. In the case of our database, there would be a single object (Images) that contains a number of different variables (inventor name, author name, issue, date published etc.). To put content on the database, the .xml file that contains the actual information must specify some or all of these attributes for each individual entry. The templates are HTML files with dynamic elements to do things like display more than one result at a time. The HTML files describe what the site will look like and how information is ordered. Views contain information retrieved from the database in the particular ways used by the templates. The framework combines all of these background components into a website.

One of the biggest decisions was what information to include for each image and how to structure it. The layout and functionality of the site is simple and all the information we want to record is relatively distinct; it made sense to give all the information about an image its own field rather than sub-organizing it as the previous group did. Putting all that information into distinct fields was relatively simple.

Django XML format was challenging to figure out, but it allowed for document creation by filling the database via the browser form and dumping it into a file. With knowledge of database structure, a program was written to convert our format so that the information could be entered into the main database and used. This was done in Java and involved two steps: enter all

the information into custom data structures then print those structures out in the required format. This two step process simplified the conversion of the old database as well as our custom format. Both needed their own code to record all the information, but since the type of information was the same, they could both be readily put in XML format.

Although it was working in a test environment, deploying our database proved difficult. In the interests of flexibility and usage dynamics, Django requires a continuously running process on the server hosting the website. Our site does not require this; one search returns a series of result and all future searches are independent. The only real information the site requires is that specified by the user. The site need only fulfill requests to generate and send back a simple HTML page. A continuous process framework program is high demand, so it is considerably difficult to find limited budget hosting for a site like Django. Ideally our database could easily be run anywhere, such as on WPI resources. Without a simple means to put the site online with Django, it made sense to find another way.

Final Implementation

The simplest alternative was to create and fill a database then put a file in a place accessible to the Internet that would search through the images on the database and organize that information into a page. The biggest reason for this approach was the straightforwardness of implementing it on WPI resources; there are relatively simple ways of creating databases and putting things online. The drawbacks were increased complexity--the need to make requests for information to the database directly, and to organize all page transitions with a single file--and the need to manually program many of the things Django already provides, such as organizing results by date and filling in the database via an online form. Though working with the database directly was more laborious, it was easier to figure out. Django, being a framework, had various

built-in functionalities that had to be learned to be used. A MySQL database with a CGI script, on the other hand, could be made to do the same things through elaborate combinations of basic programming techniques. For instance, in its current form the site decides whether to display the basic or advanced search fields by looking in the URL--this is 'form data', which technically doesn't need to be visible to the user. If in a certain place it sees the text '=true', it displays the advanced fields. If it instead sees '=false', it shows the basic search bar.

The essential structure of our current implementation is a MySQL database, a CGI script written in Python, and a series of programs to convert the text files written to index the material into the database. The MySQL database is composed of a single table, titled "images," containing a number of different word and number variables. A table is like a pile of identical forms, each representing an image, with some or all of the information on them filled out. Running a query on the database requests forms containing specific information, such as forms where the page number is greater than seven, or forms where the image description contains a certain phrase. The database then sends the matching information to the program that queried it. The CGI script creates the webpage itself. When you go to it from a browser, it retrieves the form data, checks if searches are specified, and, if so, queries the database regarding these searches. The browser-readable HTML code to display the text and the search fields is then printed, followed by more HTML describing all the search results and other features on the page. These results are initially uploaded into the database with programs that read text files. In general, this is done by retrieving information piece by piece, and once all the information for an image is collected, the program connects to the database and puts it in.

While some of this was different enough from the original system that it had to be done from scratch, much of it carried over. The structure of the templates and some of the HTML code

in them could be used. The program to convert the old database into Django XML, along with the program to convert ours into the same, could be temporarily useful as part of a two step process involving conversion into XML format followed by processing into the database. This was, in retrospect, not the most efficient way to go about it; any future entries into the database would take two steps using first the Java program used to parse our format into XML then the program to take XML and make it into a series of table entries. In the interest of simplifying the system, another Python program was made to take a text file in our format and enter it into the database directly, as well as a program that takes everything on the database and makes a file in our format.

One of the things the old group considered as a potentially useful feature was a web based data entry form. While we did not implement it either, it would be relatively simple, especially with the existing entry program. A CGI script could take a long piece of form data, entered into the page, and run the exact same code as would have been run on a text file on the page data to put it into the database. However, the utility of this feature is debatable--all entered information can be easily transferred to the database as is. If it was done this way, some sort of password and possibly user system would have to be set up to control who could add information to the database, and that would be subject to security concerns. As is, it would not be very difficult to enter form data in the URL manually that could alter the contents of the information in the database, and it would make sense in the future to screen that data for query related keywords and related symbols.

Something we strongly considered but did not have time to implement was adding URLs linking to the actual images reference. This is as simple as filling in the URL field for any given image entry in the database, with the only complicated part being finding the URL for every

entry. This could be accomplished by writing a click bot to go through the site with the images (http://cdl.library.cornell.edu/moa/moa_browse.html), find each one, and manually copy the URL to transfer into the database. The most straightforward way to do this is to write a Java program utilizing Java.Robot; it would go through the database and extract information on the title, year, date, volume and issue of each image. The program would then iterate through them, and for each, it would use Java.Robot in conjunction with clipboard information to navigate to the page with the image. That is, on the page linked, it would go to the appropriate link by refreshing the page (F5), pressing tab a number of times relative to the date of the current image (ie., the number of times to get to the first date link, plus the year of the image in question minus 1846), and hitting enter. To determine which volume/issue to click on, you would need to have Java.Robot press ctrl-a and ctrl-c, extract the clipboard information and parse it into a list of all the links, and further parse out the volume and issue number from those. Then simply iterate through the list, looking for a link with volume and issue numbers that match that of the current image, and press tab a number of times relative to the index of that entry in the list to select the right link. Repeat this process for the page, and you will be at the image. At this point, the program would have to select and copy the URL, and use the title of the image to insert that URL into the database at the entry for that image. To reset the page for the next image, have Java.Robot press backspace three times. It might be helpful to have delays between commands (`r.delay(300)`), to compensate for slight delays in page loading. The code for getting the clipboard information is as follows:

```
static String returnClipboard()
{
    String[] strings;
    String s="";
```



```

        Clipboard clipboard =
Toolkit.getDefaultToolkit().getSystemClipboard();

        Transferable t=clipboard.getContents(null);

        try
        {
            s=(String)t.getTransferData(DataFlavor.stringFlavor);
        }

        catch(UnsupportedFlavorException ex)
        {
            System.out.println("...");
        }

        catch(IOException i)
        {
            System.out.println("...");
        }

        //s=
        return s;
    }

```

How it works

To give a detailed description of the inner workings of the CGI script, the following is a step by step process of how it works when the program is run:

1. A variable is created that holds a connection to the database made with MySQLdb.
- 2. A variable to manage form data is created, a string (segment of text) is created that is the same as the end of the URL containing the search information. Later a link can be generated for next and previous pages that will have the same information.
- 3. Various information is taken from the form data, including everything that may have been

typed into the relevant search fields.

- 4. A query string is generated. For every search term that was retrieved from the form data, more information is added to this string, making the search more specific.
- 4b. If the form data says that the current page is intended to display information on an image, the query string is replaced by a search for images with a specific title (intended to give only one result, with all the information about that image).
- 5. If there was any date information in the forms, the program checks if the month and day are two letters long. If they are not (if it was the fifth day of the month for example), then it puts a zero in front of them. This is because the next thing it does is combines them all together into a single number, organized from year to day, for the purpose of organizing by dates. For this to work, if one date occurred later than another, it has to generate a higher number, which is ensured by checking that the physical length of each part (month/day/year) is the same length. So Feb 5, 1875 would become 18750205, which is larger than Feb 1(18750205), which is larger than January 20(18750120). There is a similar system in the program that records information into the database to create a field in the images table with number versions of dates, so images can be queried by these. A query is generated from these numbers and added to the general query string.
- 6. The query is run, returning a variable that can be used to retrieve results. Based on the page selected, the program 'scrolls' forward on this variable, so that when you go to retrieve results, they will start further along (if it is the second page, it scrolls (number of results per page)*2, etc.) A list is filled with each result after that up to the number of results per page, taken from the query made on the database.
- 7a. If the form data indicates that the current page is not intended to display information

about an image (ie. If there is no information describing what the title of the image is in the form data) then the HTML code representing the basic search bar is printed (printing something in a Python CGI script will cause the browser to display it like an HTML file). If the form data indicates that advanced search is selected, it prints an HTML link back to basic and additional search fields. It then prints links generated from all the search results, and links to the previous/next pages.

- 7b. If on the other hand, the form data indicates a display page, the program prints out a link back to the main page of the site and all the information on the first image in the search result list. Because the query was changed to be for a single image, this will give the image specified in the form data.

7. Future Work

Although several issues of *Scientific American* have been added to the database, there are fourteen more volumes available on Cornell University's site. Additionally, there are many other periodicals available on their Making of America site that should someday be cataloged, including *Harper's New Monthly Magazine* and *The North American Review*. Furthermore, the American Antiquarian Society has many periodicals in their collections that could be cataloged and even possibly made available online.

Some improvements could be made in the *Scientific American* database. Providing a web link to the actual image would be very beneficial to the user, as well as including the actual image. An independent website outside of Worcester Polytechnic Institute's web server would greatly aid the accessibility of the database. Also, if the cataloging files could be directly uploaded to the site without use of special software, more individuals could add to the database with ease.

8. Conclusions

This database offers a wealth of new information for the academic community. A simple text query brings up related images from *Scientific American* prints. These images, often lifted right from patent submissions, give the researcher a deeper level of understanding of the invention than just text. Easily accessible image data will aid not only the history scholar but those interested in art, technology, journalism, and other subjects.

Our online database, powered by a CGI script reading MySQLdb, is designed to promote exploration and revision of the cataloged information. Intuitive categorization of data and fairly standard digital formatting ensure that future groups can expand upon our work. Some additional features such as embedded images and online entry submission could be added to improve the interface.

Scientific American and Munn & Co. helped encourage innovation in America from the 1850s into the 20th century. The widely-accessible and distributed publication, supplemented by financial and legal support for rising inventors, made an undeniable impact on technological advancement. Now, this fact can be better appreciated and explored through the use of our database.

9. Citations

"A History of the Patent Law of the United States." Intellectual Property Law. 12 Jul 2003.

Ladas & Perry. 11 Feb 2009

<<http://www.ladas.com/Patents/USPatentHistory.html#fnB12>>.

"About Us: History." Scientific American. Scientific American. 11 Feb 2009

<<http://www.sciam.com/page.cfm?section=aboutus>>.

Bellis, Mary. "The 212th Anniversary of the First American Patent Act." About.com: Inventors.

About. 12 Feb 2009 <<http://inventors.about.com/library/weekly/aa073100a.htm>>.

Dobyns, Kenneth W. The Patent Office Pony: A History of the Early Patent Office (1994)

<<http://www.myoutbox.net/popstart.htm>>.

Evelyn, Douglas E. . "Exhibiting America: The Patent Office as Cultural Artifact." Chicago

Journals (1989) 24 Feb 2009 <<http://www.jstor.org/stable/3109008>>.

Khan, B. Zorina. "'Not for Ornament': Patenting Activity by Nineteenth-Century Women

Inventors." Journal of Interdisciplinary History 31. 2. Autumn 2000 159-195. 17 Feb

2009 <<http://www.jstor.org/stable/pdfplus/207141.pdf>>.

Khan, B. Zorina. "Property Rights and Patent Litigation in Early Nineteenth-Century America."

The Journal of Economic History 55. 1 March 1995 58-97. 17 Feb 2009

<<http://wf2dnvr13.webfeat.org/1cTZL1222/url=http://www.jstor.org/stable/pdfplus/2123768.pdf>>.

Jones, Stacy V.. The Patent Office. New York: Praeger, 1971.

"Patents: The Collection for All Reasons." United States Patent and Trademark Office. United State Patent and Trademark Office. 17 Feb 2009

<<http://www.uspto.gov/web/offices/ac/ido/ptdl/patreaso.htm>>.

"Orson Desaix Munn." Appletons. 2001. Virtuology. 14 Mar 2009

<<http://www.famousamericans.net/orsondesaixmunn/>>.