Worcester Polytechnic Institute
Digital WPI

Interactive Qualifying Projects (All Years)   Interactive Qualifying Projects

May 2008

Images in Mid-Nineteenth Century American Scientific Periodicals

Allison P. Smyth
Worcester Polytechnic Institute

Jonathan David Gibbons
Worcester Polytechnic Institute

Nicole Marie Nelson
Worcester Polytechnic Institute

Stephanie Jenna Fuller
Worcester Polytechnic Institute

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

Images in Mid-Nineteenth Century American Scientific Periodicals

An Interactive Qualifying Project Report
Submitted to the Faculty
of the
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirements for the
Degree of Bachelor of Science

Submitted by:

Stephanie Fuller
sfuller@wpi.edu

Jonathan Gibbons
jonored@WPI.EDU

Nicole M. Nelson
nmnelson@wpi.edu

Allison Smyth
asmyth@wpi.edu

April 24, 2008

Abstract

This project, sponsored by the American Antiquarian Society, created a database of information about illustrations in early volumes of *Scientific American* (1846-1854). Electronic indices of historical documents usually allow searching only for text rather than illustrations. This database, built on RDF and accessible through a web interface, provides searching for structured, textual descriptions of illustrations. The report discusses ways of extending the project as well as the historical context of *Scientific American* and the technology of its time.

Authorship

Edited by Stephanie Fuller and Jonathan Gibbons

Contents

Chapter 1

Introduction

In the 2007-2008 academic year, our group conducted a project in which images from early American scientific periodicals were cataloged. The main purpose of the project was to create an image database that would improve the process of searching for scientific illustrations in nineteenth-century periodicals. Largely due to the material that was found when we began cataloging illustrations, a portion of the project work came to be dedicated to studying the role of technology in American society in the years between 1846 and 1854. Although the main focus of the project continued to be the construction of an image database, it was hoped that additional research would lead our group to better understand the context of the images that were cataloged. It was also hoped that the research would possibly lead to a better understanding of the general history of technology as well.

Our project began when Gigi Barnhill of the American Antiquarian Society (AAS) recognized a problem that many scholars faced when conducting research in periodicals. In her work, Ms. Barnhill often noticed the difficulty of locating illustrations in early American periodicals. While the texts of these periodicals are becoming increasingly easy to search, the images within these periodicals remain relatively inaccessible. Finding images of a certain genre or a specific subject is generally a tedious and difficult task. Therefore, to aid historical research, Ms. Barnhill asked for a database listing the images from nineteenth-century American scientific periodicals held by the AAS.

At the current time, work has been initiated by scholars in order to improve access to early American periodicals. The pinnacle of this work is the Making of America Project, which provides a digital library of nineteenth-century publi-

cations that includes a number of periodicals. This library makes viewing early American periodicals easier, and further increases accessibility to the content of the periodicals by converting the page images of the documents to text. This process converts the information from the periodicals into a searchable format; however, it does not aid art historians or other researchers who wish to examine the illustrations.

A database of images would require significantly more work to create than a similar database of text. Unlike text, it is difficult to search for images in an online collection since they cannot be located by simply typing the desired entity into a search bar (one cannot insert a picture into a search bar and expect the computer to find the desired image). Therefore, a system would have to be developed in order to relate the pictures to words so that it would be possible to find the desired picture based on a description.

This project created a foundation for a database of illustrations from nineteenth-century American periodicals. It was decided that a single periodical from the mid-nineteenth century should be the focus of the project. Our group chose to index *Scientific American*, a mid-nineteenth century scientific periodical that is now the oldest continuously published magazine in the United States.

Our group also needed to determine a scheme for cataloging the images of the periodicals. This included deciding what information should be cataloged from each image as well as how this information should be entered into a database. The final project included creating an online user interface with the database in order to make the research conducted and the images cataloged available to researchers.

Currently, the illustrations from volumes 2 through 9 of *Scientific American* have been cataloged and inserted into the database. The database was designed to be easy to edit, which will allow additions to be made to the database without extensive programming. A general format for cataloging images was created as the images were cataloged from *Scientific American*. This format was also reviewed by Ms. Barnhill who offered suggestions about what information would be necessary for historians and academic researchers who would be using the database. Although only *Scientific American* was cataloged during this project, the format for cataloging images was designed so that it would apply to all periodicals cataloged in the future.

Although a large portion of the work conducted involved cataloging illustrations from *Scientific American*, research was also an important aspect of this project. Our group met with librarians from Webster, Massachusetts, as well

as WPI's Gordon Library in order to learn about existing work in the field of cataloging. The purpose of this research was to make our database easier to use for researchers by adapting our research in *Scientific American* to the searching techniques that scholars are accustomed to using.

To analyze the illustrations in *Scientific American*, an examination of history of the mid-nineteenth century was conducted. This research allowed us to better understand the context of the illustrations found within *Scientific American*. Finally, research was conducted regarding the changes noticed within the various volumes of *Scientific American* in order to show how *Scientific American* evolved over the time period reviewed.

Although the database is not complete, it was understood at the beginning of the project that this would not be a feasible goal. Instead, the project created the foundation for an image database that could be expanded in the future. The following sections review the cataloging that we accomplished as well as the research that we have conducted throughout the project. It is hoped that the cataloging and research will help future groups continue the project and eventually create a valuable tool for researchers interested in nineteenth-century American technology.

Chapter 2

Our Sponsor

The sponsor for this project is the American Antiquarian Society (AAS), an independent research library at 185 Salisbury Street in Worcester, Massachusetts. The library was founded in 1812 by the printer Isaiah Thomas. Thomas's goal in establishing the AAS was to "encourage the collection and preservation of the Antiquities of our country" as well as to "collect, organize, and preserve the records of the lives and activities of people who have inhabited this continent" in order to "encourage the study and understanding of the past" (McCorison and Hench).

The AAS is a national center for research with an extraordinary collection of early American materials including nineteenth-century periodicals. It is primarily concerned with collecting and making available printed materials that were published prior to 1876. Although it is often considered a museum of American history due to the large collections that it preserves, the AAS is an essential resource for researchers interested in the history of America.

At its present location, the AAS stores the entire library's collections, which document the life of America's people from the colonial era through the Civil War and Reconstruction. Collections include books, pamphlets, newspapers, periodicals, broadsides, manuscripts, music, graphic arts, and local histories. The AAS attempts to make history more accessible to researchers and scholars through observing and preserving the relics of the country. It also aims to better understand the history of the United States through research conducted on the items in the archives.

Cataloging images from nineteenth-century periodicals was of particular interest to the AAS in order to make early American illustrations more available

to researchers as well as to encourage a deeper understanding of the history of technology in the United States during the mid-nineteenth century. This time period in America was characterized by rapid industrialization and the onset of modern technology; therefore, the AAS hoped that through increasing the availability of scientific illustrations from this period, the period itself would be better understood.

Chapter 3

Cataloging

Cataloging, defined as creating a description of an object to be cross referenced and searchable, is a major field of study in libraries, museums, and information technology. These disciplines in particular have a strong tradition of cataloging as they are primarily concerned with the management of information. In order to avoid having to learn each of a multitude of cataloging systems created by diverse groups, cataloging standards have been created.

3.1   Common Catalog Standards

There are several different styles of cataloging used today. Most catalogs follow the rules of the International Standard Bibliographic Description (ISBD) that was created by the International Federation of Library Associations (IFLA). One of the most commonly used rules for cataloging are the Anglo-American Cataloging Rules second edition (AACR2), which provides rules only for descriptive cataloging. A similar set of rules to the AACR2 are the Library of Congress Rule Interpretations (LCRI). Dewey Decimal numbers are also used to help in the cataloging of texts. The numbers are generally assigned to U.S. trade imprints and to texts in foreign languages.

Most libraries used card catalogs to organize and help search for texts, but as technology advanced the system of cataloging also advanced. Online catalogs are more popular today than the traditional catalogs and generally follow the Machine Readable Cataloging (MARC) standards.

Dublin Core (DC) is a cataloging standard which originated in information

technology. Dublin Core is often used in information technology, and its use has expanded beyond this field. The MARC and Dublin Core standards are roughly equivalent, with formal transformations providing the ability to convert from one to the other, although Dublin Core seems a more modern standard.

3.2   Catalogs Concerning Images

The Library of Congress has a catalog that contains prints and photographs that can be accessed online called the Prints and Photographs Online Catalog (PPOC). The catalog contains descriptions of groups of images that have been cataloged since 1945 and single images that have been cataloged since 1984. Depending on the collection, not all of the images pertaining to it may have been digitized.

The PPOC can be searched by the text fields which include the author/creator fields, title fields, subject fields, and number fields. The catalog can also be searched by the collections or categories that the Library of Congress has created in the catalog.

The American Antiquarian Society also has a public catalog of images, the Catalog of American Engravings. This catalog covers engravings, either published separately or as part of a book or periodical, from early 18th century through 1820. This catalog is based on the MARC format, and include information including the size of the illustration, where it was printed, and a description of the image along with other descriptors. They also include custom subject headings rather than using those from the Library of Congress. This catalog has a Boolean search on any individual field, or all fields (Barnhill).

3.3   Our Cataloging

This project used a cataloging scheme related to Dublin Core, but with more detailed information on several fields, notably including machine-readable description of individual inventions (or other things) that the images were about. Terms used in this project can be converted to equivalent terms in Dublin Core by computer, for cases where the information is expressible in Dublin Core. Details on the terms used in this project are presented in section 4.3.

Chapter 4

Our Work

4.1   Project Goals

The purpose of this project was to create a database of Early American scientific illustrations for the AAS while learning about the interaction between technology and society. Since we decided to focus the cataloging on *Scientific American*, we wanted to extend over as large of a time period as possible starting with the first volume. Along with the cataloging, we were to create a database which could have other periodicals added as the project expanded beyond our work. Along with this we were to create a web interface which allowed historians and other researchers to easily find illustrations they are looking for or look for trends of what was pictured. The cataloging, creation of a database, and creation of a web interface were all necessary components of this project.

In order to best fulfill what the AAS wanted, we created some of our own requirements. One of the more important ones for continuing work outside of our project was the ability to have new information easily added by non-technical people.

In general this project was supposed to improve accessability of early American scientific illustrations by creating a database of these illustrations.

4.2 Related Work

4.2.1 Structured Query Language

Invented in 1974, the SQL standard for relational databases is one of the defining characteristics of its field. It allows the straightforward manipulation of tables of data, which are arranged according to an entity-relation model of some knowledge that is to be queried. Important aspects of this are that each entity or relation has a fixed set of fields that are associated with it. This allows substantial optimization in many aspects of the database, and helps to ensure that the internal consistency of the database is maintained. For our applications, however, internal consistency is of relatively little concern, and the optimizations are not generally applicable to the task at hand. This means a less constrained form is preferable.

4.2.2 Resource Description Framework

RDF defines an abstract model of metadata, defined as data about data[1], and a set of mechanisms whereby that metadata can be communicated in a machine-manipulable manner (Klyne, Carroll, and McBride). It also presents a manner in which new properties can be defined, and optionally declared as refinements of older properties.

RDF metadata is defined for resources, which are simply the name that RDF gives to any object that has been described in the framework.

The abstract model of RDF represents information as a set of statements of the form "subject predicate object," where the predicate specifies what the relationship between the subject and the object is. For instance, a viable statement is " 'illustration 1' has-subject 'turbines' ." Identity is preserved for the elements in these tuples. Two statements about the same illustration have as a subject something that can be determined to be the exact same object.

Key to the operation and interoperation of systems based on RDF is the idea of namespaces. If there were no method to assign a unique name to each predicate, then one application's idea of a "subject" may very well be different from a different application's idea. Namespaces resolve this issue by associating a Universal Resource Identifier[2] with each set of names, and so one applica-

---

[1] For example, a Library of Congress subject heading in regards to a book. It is information about the information contained in the book.

[2] Essentially, a URL - http://example.org/ and the like, but there are some technical nuances of little importance here.

tion's "subject" is associated with one URI, and the other's is associated with a different one. Generally, these URIs are also methods to discover the formal ontologies associated with those terms. This essentially establishes a controlled form of identity for terms.

RDF also defines several representations for data; one of them is based on the Extensible Markup Language and intended primarily for computer manipulation, and the other, called Notation 3, is intended as a more compact notation for human entry and reading. It is worth note that Notation 3 is essentially a more developed version of the notesheets we had been using for text entry of data.

SPARQL Protocol And RDF Query Language

Along with the RDF model, there is a defined standard query language. This allows information to be retrieved from the database where some pattern of tuples matches (Prud'hommeaux and Seaborne). The patterns are allowed to contain variables, which are constrained to represent the same object, in the same manner as formal logic and logical programming languages such as Prolog. The information retrieved is essentially a list of bindings for these variables, where the bindings are a list of objects from the store that fit the pattern.

Notation 3

Notation 3 is a format for writing RDF files by hand; an example is given in Figure 4.1. Key points are the relative simplicity of each statement, and the ability to include terms from multiple vocabularies (Berners-Lee).

The coexistence of multiple sets of terms in one file or store is one of the strong points of RDF; in N3 it is accomplished by way of using full URIs for each term, and having an abbreviation mechanism in place by letting prefixes[3] be defined in a method similar to XML Namespaces. In Figure 4.1 the two subject headings "power" and "waterwheel" are declared using the 'subject' term of the DCMI Metadata Terms, which is defined to be used for the topic of the resource, while 'subject' in this project's namespace is used for the specific entity that the image concerns, in this case, Mr. Sherrod's water wheel.

Another useful aspect are the various abbreviation mechanisms for statements. In RDF, a statement is always of the form "subject term object." How-

---

[3]The line at the top of the example is such a prefix declaration; similar declarations for the el: and q: prefixes are defined at the top of the overall file.

```
@PREFIX dc: <http://purl.org/dc/terms/>
el:a101
 q:title "Sherrod's Water Wheel Plan from Front";
 q:description "This illustration rather than giving a view of what Sherrod's
                water wheel looks like, gives a design for making this type
                of water wheel. It mostly consists of labeled lines with
                descriptions in the text. ";
 q:keyword "design",
           "fan water wheel",
           "plan",
           "power",
           "water wheel";
 q:published [ q:col "2";
               q:date "November 6, 1847";
               q:issue "7";
               q:page "49";
               q:publication "sciam";
               q:volume "3". ];
 q:type typ:illustration;
 dc:subject "power";
 dc:subject "waterwheel".
el:a101 q:subject el:s50.
el:s50 q:description "A horizontal water wheel in which the paddles are
                      turned by the current of the river. The idea behind it
                      is that changes in the river height, or even freezing
                      of the river, would not affect the water wheel.";
       q:genre "mechanical";
       q:creator "Mr. W. Sherrod";
       q:title "Sherrod's Fan Water Wheel";
       q:type "subject".

el:a101 q:subject el:s51.
el:s51 q:title "fan water wheel";
       q:type typ:subject.
el:a101 q:subject el:a24.
```

Figure 4.1: Example of Notation 3

ever often a series of statements with the same subject needs to be added; or even a sequence with the same subject and term. In Notation 3 if a semicolon is used instead of a period at the end of a statement, the subject of the next statement is omitted and taken to be the subject of the preceeding statement. If a comma is used, both the subject and the term are. In the example, statements are generally chained with semicolons and keywords are listed with the comma notation.

The other feature of note is the use of a blank node for the publication event. This event is not really such that it should have its own name, because it is only interesting as a relationship between the publication that the illustration was printed in and the illustration. The blank node notation has the form "subject term [term object; term object; . . . term object .]." The example can be mapped into english as "[the waterwheel] has relation 'published' to the event of being published in volume three issue seven of <u>Scientific American</u> on page 49 column 2, on the sixth of November, 1847."

RDF Schemas and the Web Ontology Language

To assist in verification and computer manipulation of RDF stores, a schema system is also defined by the W3C (Brickley, Guha, and McBride). This schema system allows the machine-readable description of a vocabulary of terms, including definitions of acceptable subjects and objects for properties, and basic interrelations between vocabularies. The Web Ontology Language extends RDF schemas with many ideas from formal logic (Bechhofer, van Harmelen, Hendler, Horrocks, McGuinness, Patel-Schneider, and Stein).

4.2.3 Dublin Core

The Dublin Core Metadata Initiative is a standards group started in Dublin, Idaho to provide a standard base set of metadata attributes that are applicable to general resources (DCMI). The Dublin Core community contributed to the activity leading to RDF, and Dublin Core often provides a sort of lingua franca for RDF when applied to catalogs. Because of RDF Schemas new properties can be defined as sub-properties of DC terms, providing a mechanism for describing specialized local versions of terms while maintaining general compatibility. Dublin Core is used widely in cataloging systems, as metadata embedded in web pages, and as the supported format required of implementations of the Open Archives Initiative's interoperability protocols. It consists of fifteen broad base

terms, spanning the majority of the information one would want to know about a resource. The terms are reproduced in Table 4.1. The Dublin Core has been recognized by the International Standards Organization as ISO Standard 15836. In addition to the base fifteen elements, the DCMI also provides a set of more specific elements, called Qualified Dublin Core (DCMI-Terms).

Element	Description
Language	A language of the resource.
Subject	The topic of the resource.
Source	A related resource from which the described resource is derived.
Rights	Information about rights held in and over the resource.
Publisher	An entity responsible for making the resource available.
Type	The nature or genre of the resource.
Title	A name given to the resource.
Creator	An entity primarily responsible for making the resource.
Description	An account of the resource.
Relation	A related resource.
Date	A point or period of time associated with an event in the lifecycle of the resource.
Contributor	An entity responsible for making contributions to the resource.
Format	The file format, physical medium, or dimensions of the resource.
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
Identifier	An unambiguous reference to the resource within a given context.

Table 4.1: The Dublin Core Element Set

4.3 Cataloging in this Project

4.3.1 The Information we Tracked

The domain covered in this project contains several different classes of interrelated objects; most straightforward is the illustration itself, which is the focus of the project, but several other entities are also involved. The people who worked on illustrations are another, as are the subjects of an illustration. The

event of a publication is also a separate entity from the illustration, becuase an illustration could be reprinted multiple times in different places (and was, in the case of many of the advertisements in *Scientific American*). Also tracked are the series (recurring column), which many illustrations in *Scientfic Americang* were published as part of. Each of these entities had properties associated with them. As these properties are described, each will be given an example using the illustration in figure 4.2.

The first property is the image title. If the illustration has a title in the article this is the title used and if it doesn't then we often used the title of the article (followed by which figure if there was more than one). For the example illustration the title is "Sherrod's Fan Water Wheel" as appears in the image.

The illustration also has a description. This description we used to explain information about the image itself - not about the thing pictured. This includes information such as the view of the thing pictured along with other details about the illustration itself. Sherrod's Fan Water Wheel has the description "This illustration shows what the fan water wheel Sherrod made would look like from the side."

Theoretically we were also cataloging the medium (wood-cut, copper engraving etc.), though this was not practically being tracked. While having the medium would be useful for historians looking for patterns in printing methods, all of *Scientific American*'s illustrations were wood-cuts, and this rule can be applied automatically by a computer.

Another property we were cataloging were keywords. In our cataloging keywords could be of either the illustration or the subject (as defined below), without destinction. [4] This was one of the most important properties because of the relationships between illustrations that keywords would allow searching on. An example list of keywords is "water wheel, mechanical, power, fan water wheel, water, hydraulic, motion."

The places the illustrations appeared were also cataloged. This information began with publication, though we were working within one periodical. For publication we used 'sciam' as an identifier for *Scientific American* - the full title of *Scientific American* is presented to users of the database, but a shortened form helped with cataloging. Unlike the periodical, tracking the volume, issue and page were straightforward and necessary even within this project. Also, the column on the page was tracked. Because there were illustrations which spanned

---

[4]This destinction may be good in the future. Also a possibility is if keywords of the illustration are not used, making keywords a property of the subject.

more than one column, the leftmost column was the one stored. Though the information on the date of publication of an illustration is associated with the issue, the date for the illustration was also directly tracked. This is mostly so historians could more easily look for changes over time. Our example illustration was in 'sciam,' volume 3, issue 3, page 17, and column 2. It appeared on the issue from October 9, 1847.

Similarly to where the illustrations appeared, any series they were published as a part of was also kept track of. The name of the series was cataloged as well as a description for the series. This description explained what the series was about and the purpose of the series, but not information about the illustration or the item pictured. If a illustration was not in any series this was left blank. Our example illustration did not appear in any series, so this property did not exist for it.

We also kept track of the information we could gather about the people involved in the publishing process. Though most illustrations in *Scientific American* did not have any information on the artist, engraver, or printer, there was some information we were able to gather. In this section there were two properties we were cataloging - the name of the person who worked on the illustration, and the task they completed. These two properties were separate because there were situations where we had a name and no idea what this person did on this illustration. In this situation the person's name was given and the task property was left empty. If no information was given, such as in the case of Sherrod's Fan Water Wheel, the properties were left blank.

The thing pictured was also cataloged, under the property subject. Every item had a subject, and the same subject was listed for all illustrations of the same thing. As with the illustration, the first property is the subject name. To include more detail there was a subject description which explained what the subject itself was. This often included information found in the articles which the pictures themselves did not portray. However, if the subject was self explanatory the subject description could be omitted. The genre of the subject was also tracked - whether it was a mechanical device, electrical device, architectural item, etc.[5] For the items we knew who created it this was also listed under creator. Most of these were inventors however there were times

---

[5]Genre was an early addition to what we were cataloging, which looking back was trying to accomplish some of the way we multiple subjects accomplishes. Because whether something is mechanical, electrical, etc, is important, we wanted to track this information. Because we hadn't realized the appropriate thing would be having a set of mechanical objects with the subject mechnical, we tried to make up for this using genre.
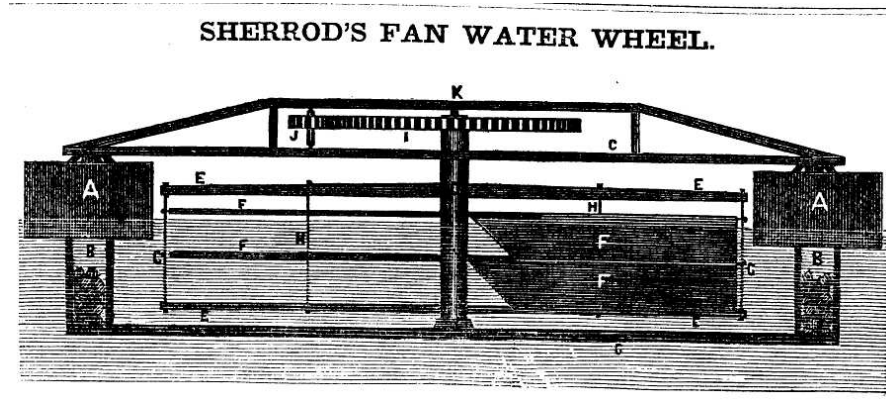
Figure 4.2: Example from *Scientific American*, Vol.3, No.3[1847]

when the item pictured had not been invented. Illustrations also, had multiple subjects applied to them. This is discussed in more detail in section 4.3.3.

The illustration of Sherrod's fan water wheel, gives examples of multiple subjects. The first subject was, like the illustration title, "Sherrod's Fan Water Wheel."[6] This subject came with the description, "A horizontal water wheel in which the paddles are turned by the current of the river. The idea behind it is that changes in the river height, or even freezing of the river, would not affect the water wheel." The creator was listed as the name given in the article, Mr. W. Sherrod, and the genre was mechanical. Along with this subject, which refers to the specific item pictured, there were also the subjects of 'fan water wheel,' 'waterwheel,' and 'power.'

4.3.2   Optional Properties

In general properties which did not make sense to have for a specific illustration were not included.If there was no information about who worked on the illustration, the information was not simply not included. Similarly if the illustration did not appear in a series, then the 'series' property was left blank. However, we decided not to include a task someone had on an illustration without information on who this person doing that task was. As we know people did draw,

---

[6]While in this case the illustration title and the 'specific subject' are the same, this is not necessary. To give an example, in volume 3 issue 3, there was an illustration with title " Electrotype and Electro-Gliding - Figure 1 " (*Scientific American* Vol. 3). The subjects connected to this illustration are 'Simple Battery' and 'battery.'

engrave, and print the illustrations, so there is no need to say that these tasks were completed without more detailed information. Without more information the fact an illustration we see has been printed is not historically interesting and including this fact would add nothing of value. Similarly, we could not have a series description without a series.

Subjects can also be included and left out as a total entity as needed. Similarly, the subject description was left out when there was no description to add to the subject title (such as with the subject 'waterwheel'). Generally forcing information which did not make sense or did not add any information was simply not done.

4.3.3 Multiple Subjects

Illustrations in general had multiple labels which could be given of what was pictured. An illustration may be of a waterwheel and a boat, in which case, both 'waterwheel' and 'boat' describe what is pictured. On the other hand, there may be an illustration of 'Sherrod's Fan Water Wheel,' which may also be labeled as the less descriptive subject 'waterwheel.' In both these cases, this is dealt with by the use of multiple subjects.

Part of the necessity of multiple subjects is that there are two classes of subject undifferentiated in the database. One of these classes of subject is the specific referent of the illustration. There is generally one of these per illustration, though it is possible for an illustration to have multiple or even none. One of these subjects would be 'Sherrod's Fan Water Wheel.' It is the most specific of the subjects and something like Library of Congress subject headings would not work for these.

The other class of subjects is the general subject headings. These could be, and in future projects should be, determined by a standard. Late in the project it was determined these should probably have been Library of Congress subject headings. The use of a standard allows for subjects to be the same between this database and other methods of research. Because Library of Congress is one of the most used standards for subject headings, and it would not take much extra work to use these for the subjects in this database, this is recommended for future work.

The general subject headings fall under multiple categories. There is, like the more specific subjects, the subjects which answer 'what is this?' These subjects such as 'waterwheel' make up a majority of the subjects. Subjects also were

used to answer the question 'what is this for?' The subject corresponding to our example illustration answering this is 'power.' There are also some subjects which do not have a specific subject that they answer. On an illustration of a geometric figure, the subject 'math' does not answer any specific question, but still gives more information about what the illustration is of.

As the cataloging progresses, new common subjects should be added to the database as they present themselves. This often means going back and amending the subjects for already cataloged illustrations to include the new subjects. This may span beyond a single project. For example, an illustration in *Scientific American*, 'Amant's Escapement,' should have the subject 'dead-beat escapement' added if a large amount on clocks and escapements are cataloged in the future. As it stands there is little justification for adding this general subject. Amending to add more subjects should be as simple as a search and insert. Editing the past work as cataloging continues gives a more complete set of subjects throughout the data.

Multiple subjects allow for more detailed sets of illustrations than keywords because of the other information involved. While two illustrations can share a keyword, this keyword does not include other information. The subject description and other associated properties do include more information about the sets of illustrations. As it was discovered in the second issue cataloged by one member, multiple subjects add enough to the project that they are effectively a necessity.

4.3.4   How cataloging was done

Multiple ways to store the data while cataloging came up over the course of the project.

One approach to storing data while cataloging was the spreadsheet program Excel, produced by Microsoft Inc. Excel has the advantage of being familiar, however is not the optimal choice.

For some of the project, some members were storing information in a table in Excel. This allowed them to use a computer program they were familiar with, and visualize a table of data. It also put the data in a form where syntax errors could not create a problem in loading the data into the database. However, a single table cannot properly represent a multiple-entity relational database. The number of subjects per illustration, or any other property with a many to many relationship, had to be capped, as there was no other place to store this

data. If an illustration should have more subjects than this cap, the number of columns dealing with subjects had to be increased, or some of the theoretical subjects would have to be excluded. A single table also forces the repetition of data which creates a longer time cataloging and necessarily more processing before loading the information to the database. Because the database should not have repetition of data, information in a single table cannot be directly loaded into the database.

In response to these problems with a single table, the data can be put into multiple tables in separate Excel worksheets, as it was during the project. Using Excel still gives some amount of familiarity with the program, though the data is not as easy to visualize in separate tables. Similarly, multiple tables does not create an issue with syntax errors. Using multiple tables fixes the major problem of being able to represent the data. Many to many relationships do not have the same limit, and data no longer needs to be repeated. Unfortunately using multiple tables like this puts you directly to the implementation level of the tables in the database. It effectively is the same as editing the raw database while in a different program. There is no other program loading the data in which would find any mistakes which you have made and constraints must be manually enforced. While this is doable, it is also possible to have a computer do some of the work that a human is doing with this method.

There are also smaller problems which are simply annoyances in dealing with Excel. An Excel worksheet cannot be read by a database and this creates another layer of work to be done. In order to load the worksheets the Excel file had to be exported to comma separated values, and loaded from there. Exporting, while not difficult, can trip people up and is awkward. While these disadvantages are not at the same level as most, they push an otherwise balanced pair of methods away from the use of Excel.

Another method used to store information during this project was in plaintext files using a template for where to put the information for each property; the template used in this project is reproduced in Figure 4.3. The template is copied into the text file, and the values for fields are written after the field name. Further illustrations simply repeat the block again in the same file.

This allows repeating blocks of the template such as the subject in order to be able to represent many-to-many relationships. Similarly, this allows dropping the properties in subject or series beyond the title in order to not repeat information. Together these have using plaintext able to represent the database without directly editing the tables. Depending on the search available, this can

```
title:
description:
keywords:
publication: sciam
volume:
date:
issue:
page:
column:
subject:
        description:
        creator:
        genre:
person:
        task:
series:
        description:
comment:
done
```

Figure 4.3: Template used to catalog illustrations

take much less time to revise than using a spreadsheet program. The group found that with the availability of general-purpose spell checking programs and search programs, this took less time to revise than the Excel tables.

However, using plaintext does not prevent syntax errors like the Excel tables do. Syntax errors could create problems with data not being loaded into the database. Working this way also requires representing the data in text, which can be harder to keep track of for some people.

A preferable approach to a newly constructed worksheet format is to use a preexisting format designed for human editing such as Notation 3 (described in section 4.2.2). Issues of computer language design are often subtle, and it is easy to create relatively major problems by apparently inoccuous decisions; and often a preexisting format has features that would be awkward to implement. The Notation 3 format provides some convenience features and a somewhat more elegant design than the format shown above while keeping essentially all of the desirable features, and should be used rather than a custom language.

Though it was not used in this project, a suggested method for future use is a web interface.[7] A web interface gives a reasonable user interface to the

---

[7]We discussed using this and decided the initial effort was not worth the advantages for

database which can fully represent the structure of the database without editing raw tables. This means details of the database can be hidden from the user without losing the direct correlation to the database. The web interface also allows the computer to actually enforce constraints. This prevents errors such as referencing an illustration that does not exist. The web interface also allows looking at the results in realtime, because it loads directly to the database. This means checking for consistency is easier to do while working. A web interface also has some of the advantages of a program such as Excel. You cannot get syntax errors, and a web form similarly is familiar to users.

However, such a web interface has its own issues. There is a noticeable amount of initial set up time for it to be created. While this will not be a problem for future use, this group had thought the skills of the group were such that this time was not worth spending. A web interface also requires either having internet access, or a local copy of the database and a way to merge the multiple copies.[8] The amount of extra effort it takes to always have web access while cataloging depends on the periodical being worked on. In the case of *Scientific American*, this was already necessary as the source for the periodical was itself online. However, a different periodical may not otherwise have this requirement.

Another good option is using an existing RDF editor; the advantages are essentially the same as the web interface, without the initial set up time or the requirement of access to a network. The use of RDF also greatly simplifies the integration of multiple datasets, as ease of such integration is one of the primary goals of the standard.

---

this group. However, as the project progressed it became evident we had made the wrong decision.

[8]This is not a novel problem - research has been done on maintaining consistent databases, though the group does not know much about this.

4.4 Technical Aspects of the Project

The final version of the project is based on an RDF tuple store, containing statements defined in our namespace. An equivalence between the terms of Dublin Core and our terms is defined, and made available to users and machine readers of the catalog. A copy of the tuple store in RDF/XML is attached, along with a copy of the web interface written in Python, and the inference rules to generate equivalent statements in Dublin Core for those properties that are reducible to each other.

Originally, an SQL database was constructed via standard methods of entity-relation modeling, but upon further examination several aspects of using an SQL database directly for this manner of task were found inappropriate. One primary concern was that the constraints imposed by the structure of an SQL database meant that the database schema had to be modified to incorporate new fields. Another was that the query language of SQL is geared towards more sophisticated queries than are involved in this project, and so the easy construction of a query from a lay user is much more difficult to implement.

4.4.1 Metadata database

The metadata database that we have produced is stored in RDF/XML, with all predicates given in our namespace. Alongside it is the map between our predicates and Dublin Core, specified in the form RDF Semantics inference rules. Specifically, this is a set of machine-readable rules that state that "if this is asserted in our database, then this other property can be asserted."

The entity-relation model of illustrations in *Scientific American* developed for the SQL implementation of the database still provides a good description of the structure of the database, even though it is no longer the only structure that can be stored in it; it is retained as appendix A.

Additions to the database by future projects can be made with a standard RDF editor, by hand, or via a web interface developed as part of the project.

4.4.2 Web interface

The web interface is built on RDFLib, a python library for manipulating RDF tuple stores. It uses a mix of the built-in functions for manipulating tuple stores and the SPARQL query engine; the query engine for finding what resources to provide information about, and the built-in functions to retrieve all information

available about the resource. A filtering mechanism is provided to allow some resources to not be displayed; for instance, the information concerning *Scientific American* as a whole should not be displayed on the report for a specific illustration.

The user interface has three methods of inputting information: a click interface, a query builder, and a search bar. The click interface allows searching by clicking on values in the display. The query builder looks like one usually encountered in library cataloging system, searching for particular strings in particular properties. The third option, aimed at those more familiar with search engines than cataloging, is an input at the top of the page that does a full-text search of property values.

The user of the click interface has merely to click on a field and a new query will be issued to the back end. The new query is the union of the current set of constraints and the new one representing that value. The current list of constraints is also available, and individual constraints may be dropped.

The query builder presents itself as a form with a drop-down menu with the predicates (fields) available in the database and a field to enter literal values. While this field can accept references to particular illustrations in the form of a URI reference, this is not intended as a general theme for most users. The default interpretation is that the tuple must have the string as a substring; a reasonable variation on the full-text search is implemented wherein a value matches if each word in the sequence is a substring of the value.

The simple search presents itself as a text input at the top of every page. Words entered here (separated by spaces), are matched against the text of the properties. The properties covered are the properties of entities of the type being searched for (usually illustration) and the properties of any entity directly linked to by those entities. This limits the use in that entities associated with the illustration but that are linked to indirectly are not searched, and is because of limitations in the SPARQL query language. Expanding the full-text search to include all property values is possible, but requires a different approach than the normal search implemented in this project, and was not completed. Any property with all words being searched for will lead to the related entity being returned. Full text search does not currently combine with the other search mechanisms - it simply searches the entire store for the terms, yet again because of a lack of time to integrate the two methods.

4.5   Project Summary

4.5.1   Summary of Completed Work

Throughout this project our group cataloged illustrations in *Scientific American* and created a web interface to a database to display this information. The cataloged volumes ranged from two through nine, with volumes three and five being incomplete. Searching for the first volume only resulted in finding the first page, resulting in the first volume not being included. The data involved in cataloging also varies between the volumes. Only volumes three and five have multiple subjects for each illustration, while it would be preferable that all cataloged data used multiple subjects.

The cataloged data was loaded into an RDF store and a database driven website was written to display the information. The website gives the ability to add and remove constraints of what illustrations are being described by clicking on links. There is also a full text search along with a query builder. The query builder searches a specific property for the substring, with the ability to specify multiple of these property, substring pairs. Information can also be viewed in terms of entities in the database which are not illustrations. This allows for information to be sorted and displayed by subject, person, or any other entity the viewer desires. Also possible is displaying all instances of a specific class, or all illustrations with an instance of that class. For example it is possible to display all series, or all illustrations in a series.

The third part of the project was research. Much of this research was to put the cataloged data into perspective. This is discussed in section 5, and gave our group a better understanding of the areas of history *Scientific American* could be classified as. There was also research into the field of study on the technical side. This allowed us to discover the proper tools for dealing with a project such as this, and is discussed in 4.2.

4.5.2   Final Work vs. Goals

While this project does not fulfill all personal goals, our group successfully created a database making it easier to find early American scientific illustrations. The database and website allow historians to search for illustrations in *Scientific American* for their research. Preparing for future groups was similarly successful. Because its built on RDF, the database can store arbitrary information as long as the information is described using RDF. The use of RDF also makes

it easy for non-technical people to load information because of the pre-existing tools.

However, the amount of cataloging we wished to complete (10 volumes) was not feasible. The estimation of time cataloging would take was low, as was the estimation of the amount of time to be spent on other parts of the project. The data cataloged similarly was not exactly what we wanted. There are differences in the data cataloged because different members of the group focused on different elements of cataloging. While it would be best if all the data included multiple subjects, this would have led to a smaller range of *Scientific American* covered. Taking into account the misestimations of time, what we completed in cataloging was reasonable, but not optimal.

4.5.3 Final Work vs. Plan

The project completed was different than the original project idea presented. While the basic project of creating a database of early American scientific illustrations remained the same, where the cataloged illustration occur changed. In our original plan, we would work from a variety of periodicals which were available at the AAS. Each member of the group was to choose a periodical and catalog as much as they could. Instead we decided it would be more valuable to get a wider range of dates rather than of periodicals and thus chose to focus on one. *Scientific American* was chosen because of interest in the topic and the fact it is available online.

Chapter 5

Understanding the work

In order to understand early *Scientific American* in context, a broad under-standing of the history of printing and of scientific publications is useful. Also useful to understand the worldview of its writers and illustrators is a look into the history of the American technology that *Scientific American* reported on.

5.1   History of Printing

The history of printing is integral to general history because of its influence on culture. Printing has existed since before 2000 B.C. and has changed through the ages (The Schoyen Collection). Gutenberg revolutionized this process around 1440 when he created a way to make movable type. The Gutenberg press allowed for large-scale production of printed material and commercialization of printing. The next large improvement in the technology of printing came in 1798 when when a machine to make paper was invented. As the price of paper dropped up to a third in 26 years, the price of printed materials dropped. In 1805 stereotyping, creating a mold from the original type to print multiple copies of the same thing without needing to reset any letters, was developed to the point of a commercial process. Improvements in the printing press led to faster printing. By 1845 when *Scientific American* was founded, printing was a widespread business (Steinberg).

The history of newspaper is relevant to printing as a whole because of their influence on culture. The first newspaper printed in America (at least the British colonies) was printed in Boston in April 1704. By the end of the year there were

three newspapers being printed between Boston and Philadelphia. At first, newspapers had not caught on, and there were none printed in New York until 1725. In 1754 only 4 newspapers were printed throughout all of New England, with the average number of copies not exceeding 600 each. However in 1775 more copies of newspapers were being printed in Worcester than all of New England twenty-one years prior. By 1810 more newspapers were being printed in America than Great Britain and Ireland, with up to 210 new newspapers between 1800 and 1810. Throughout these times newspapers also became fancier, separating advertisements out with pictures rather than having not even a line separating the advertisements from the text. *Scientific American* was created at a time newspapers were a major part of the American culture (Thomas).

Illustrations have similarly been a large part of printing since before 1500. It is estimated that about a third of the books printed before 1500 were illustrated. Movable type and wood-cuts could be combined in a press allowing simple illustration of books and periodical since Gutenberg's books.

Though periodicals were the reason illustration technology improved, *Scientific American* used the older and cheaper methods of printing illustrations. As were used in books since the 1450's, *Scientific American* used woodcuts set in the press to print illustrations.

5.2 History of Scientific Publications

Since scientific journalism began in 1665, publications have been used to communicate scientific information. As communication is necessary to pursue scientific discovery, a less personal method of sharing information between scientists than letters helped the community.

In the times of the early scientific periodical[1], the scientist was a gentleman scholar and an amateur. There was no divide between the scientists and the scholars, which led to general and literary journals including sections on science. However, there was a gap between the educated and the general population. Because of this gap, the incentives for writing scientific publications did not deal with the general public, but with the educated minority. As scientific societies were created for the scholars to meet and discuss what they did, the journals were created to have easier communication between the scholars pursuing science.

These early scientific periodicals were different from the modern magazine. Both in terms of content and the physical periodical, these scholarly periodicals were multi-volume works often closer to books-in-parts than a modern magazine or newspaper.

5.2.1 *Scientific American*

Scientific American was different than the earlier scientific journals in both the periodical itself as well as the content published therein.

On the physical side, while the earlier periodicals were book-in-parts, *Scientific American* was a newspaper (in other words, an unbound magazine). The newspaper was in general less formal than the journals, but also cheaper and easier to print. To the creators of *Scientific American*, the more expensive binding was unnecessary. Along with this, the time and money which would have been necessary to spend to create a book rather than newspaper were too great a cost. Similarly, the illustrations of *Scientific American* were often lower quality and smaller. This is because of the time and money required for the entire page high quality engravings. *Scientific American* had 2-10 new engravings every week, and did not have these resources (*Scientific American*).

The content of *Scientific American* was also different than the earlier periodicals. *Scientific American* included columns such as Mechanical Movements,

---

[1] The book this information came from focused on 1665-1790, though the scholarly scientist extended beyond that time

explained simple mechanical motions and mechanisms creating said movement. Earlier Periodicals did not have columns meant to educate the masses like this. Similarly, the focus on sharing what new had been patented was more helpful for the backyard inventor than the gentleman. In general *Scientific American* was more focused on educating and informing the average person than the scholar.

All of these differences come down to the difference in audience. While the earlier journals focused on scholars, *Scientific American* was written for the general public. Because the older journals were written for scholars, they had no need to educate the general public. To compare, *Scientific American* was created to teach farmers, families, and mechanics (*Scientific American*). Similarly, the general public did not care about the formality and could not afford the extra money. On the other hand, the high class preferred the formal journals and had the money required. *Scientific American* fulfilled a different purpose than the earlier scientific journals. They that they had different audiences, and thus they needed to be a different type of periodical.

5.2.2   The American Magazine and *Scientific American*

Scientific American focusing on the average person fit well with American magazines as a whole. Periodicals did a lot for the American. Henry Mills Alden went so far as to say "periodical literature has done more for the American people than any other" (Mott 1). *Scientific American* was one of these periodicals which influenced America and Americans. Other magazines with the same audience were created, but never lasted long. However *Scientific American* is credited with stimulating now historically famous inventors. For example, Thomas Edison walked 3 miles a week as a boy to get his copy of *Scientific American*. It is probable that *Scientific American* was a noticeable amount of why he went into inventing (Mott). He was one of many Americans who made a difference in the modern life because of this magazine focused on their needs rather than those of the upper class.

5.3 History of *Scientific American*

Scientific American was founded in 1845 by Rufus Porter. Subtitled "The Advocate of Industry and Enterprise, and Journal of Mechanical and Other Improvements," the publication was a weekly in the broadsheet format. The broadsheet format is characterized by long vertical pages, generally 22 or more inches, and is the largest of the newspaper formats. Porter himself was an inventor and after 10 months of publishing, during which the first volume was printed, he sold the periodical to Orson Desaix Munn and Alfred Ely Beach for $800. Munn was a professional publisher and founded Munn & Company with Alfred Beach. Beach was an inventor, patent lawyer, and publisher who's most famous invention was New York's first subway. Munn & Company maintained ownership of the periodical for over one hundred years, until, in 1948, it was purchased by Gerard Piel, Dennis Flanagan, and Donald Miller.

Scientific American claims to have founded the first branch of the United States Patent Agency in 1850 to provide legal advice and technical assistance to inventors. There is no clear record concerning how the agency first operated, but from what is known, it appears that the agency was similar to how patent agencies operate today. The inventor paid for advice and other services from the agency to help in the process of submitting a patent to the United States Patent Office. In 1859, a Washington D.C. branch of the Patent Agency was opened by Munn & Company along with Judge Charles Mason, who was a former commissioner of patents. By 1900 there were more than 100,000 inventions that had been patented thanks to *Scientific American*.

5.3.1 Topics of *Scientific American*

In the first issue, the owner Rufus Porter made a statement as to what the periodical would entail:

> Each number will be furnished with from two to five original Engravings, many of them elegant, and illustrative of New Inventions, Scientific Principles, and Curious Works; and will contain, in high addition to the most interesting news of passing events, general notices of progress of Mechanical and other Scientific Improvements; American and Foreign. Improvements and Inventions; Catalogues of American Patents; Scientific Essays, illustrative of the principles of the sciences of Mechanics, Chemistry, and Architecture: useful

information and instruction in various Arts and Trades; Curious Philosophical Experiments; Miscellaneous Intelligence, Music and Poetry. (*Scientific American* Vol.1)

The main topics in *Scientific American* concern new inventions. A majority of the periodical is dedicated to the discussion of the latest inventions submitted by subscribers as well as inventions directly from the patent office. Some inventions during the time of the Industrial Revolution that were described in *Scientific American* include the Bessemer steel converter, the incandescent light bulb, and the telephone.

In earlier issues mechanics and the pure sciences, such as chemistry and physics, were the main focuses in the periodical. As new, broader topics such as steam engines and electricity were created via inventions and discovery, the *Scientific American* would showcase works on the subject. Since the *Scientific American* founded the first U. S. Patent Agency branch, there was a section in the periodical that would mention some of the latest patents issued for that week.

5.3.2   Time Period of the Project

The issues covered in this project range from September of 1846 to September of 1854. The Industrial Revolution had a major effect on the sciences during this time, and this can be seen in the changing topics discussed in *Scientific American.*

In 1847, the Mexican War had ended and the United States gained the territory that is now New Mexico, Nevada, California, Utah, most of Arizona, and parts of Colorado and Wyoming. The war led to an interest in improvements in firearms and an example would be 'Hubbell's Patent Fire Arms' submitted by William W. Hubbell (*Scientific American* Vol.4, No.14).

The California Gold Rush started in January of 1848 and had an affect on the inventions seen in the *Scientific American.* There were many styles and modifications to gold washers as well as equipment for traveling west. Some of the different gold washers were the 'Gold Washer and Magnetic Separator' submitted by Samuel Gardiner Jr. and 'Barclay's Gold Washer and Amalgamator' submitted by Alexander Barclay (*Scientific American* Vol.7,No.50-51).

Steam railroads were a popular form of transportation of the time and many improvements were made to the cars and tracks. Several of the designs focused on the axles and braking systems of the railroad cars. An example of a design

regarding the axles of the car would be the 'Improved Axles for Turning Narrow Curves' submitted by Messrs. Morse and Mansfield (*Scientific American* Vol.7,No.8). An invention regarding the braking would be the 'Patent Railroad Brake' submitted by Francis A. Stevens (*Scientific American* Vol.7,No.17).

5.3.3   Characteristics of *Scientific American*

The layout for *Scientific American* changed between volumes for the first couple of years it was printed. As time progressed, the more standardized and organized the layout of the periodical became. The columns became more uniform and there were fewer changes in font throughout the periodical. The quality of the images also increased as time progressed as new techniques in printing were implemented. These new techniques also allowed more images to be shown in each issue than what could be added in the earlier volumes.

As images became easier to add in the periodical, decorative works were seen included more often. In volume five, a decorative outside border was added to the periodical whereas the earlier volume had no border. Volumes seven and eight had a different decorative border than that of the earlier volumes. Volume nine had a different decorative border than any of the previous volumes.

As the setup of the periodical changed, the opening sections changed as well. The first four volumes had a poetry section on the first column of the front page of each issue. The poetry section was replaced with a section called 'Rail-Road News' which continued into volume eight. This was then replaced with a section called 'Useful Receipts,' which was discussed new discoveries and invention such as new lubricating materials or a new type of green dye. With the start of volume nine, the magazine did away with a separate opening section (*Scientific American*).

5.4   History of American Technology

Our group also conducted research into the history of technology in the nineteenth century. The goal of this research was to provide a context for the illustrations in *Scientific American*. The mid-nineteenth century is an important period in the development of American technology. It is distinguished by the American Industrial Revolution, which was a period of rapid technological advancement that ran roughly from 1820 to 1870. It was during these years that *Scientific American* began as a means of spreading knowledge of technological

innovation.

Our study of the history of technology made the importance of cataloging images from *Scientific American* clear. The illustrations found in *Scientific American* provide powerful evidence of the evolution of technology in the mid-nineteenth century. Through studying the images, it is possible to better understand the technology of the time period. Furthermore, it is possible to see in *Scientific American* the shift from a non-technical world to one that is heavily reliant on technological innovation.

Before the nineteenth century, the term 'technology' was not commonly used in American society. In fact, according to Ruth Cowan's *A Social History of American Technology*, "no American seems to have used the term in print until the publication of Jacob Bigelow's very popular book, *Elements of Technology* in 1829" (Cowan 204). The early nineteenth century was a period of moderate technological development in America. Mechanical improvements were made gradually as they were needed. Due to the lack of mechanical developments, a term to describe technological innovation was unnecessary. By the mid-nineteenth century, the atmosphere of moderate technological advancement that Europeans and Americans were accustomed to had changed as developments were rapidly made in all scientific fields. This increase in innovation led to the term 'technology' being more commonly used.

The growth of technological innovation in the nineteenth century also resulted in the publication of periodicals about the subject. *Scientific American*, which began publication in 1845, was one of the first scientific periodicals in America. It was published with the intent of reporting events occurring at the United States Patent Office. The purpose of the patent office was to give inventors "a temporary monopoly on the economic benefit that can be derived from [their] invention[s]" (Cowan 120). It attempted "to maintain trade secrets by discussing them in public" (Cowan 123). Through these tasks, the patent office encouraged innovation in two ways: it gave inventors economic support so that they could survive with inventing as their sole occupation and it also acted to spread common knowledge. *Scientific American* furthered this goal by spreading technological knowledge throughout American society.

The following sections discuss the role *Scientific American* played in the developing technology of the period. They examine the periodical's purpose during the mid-nineteenth century while also demonstrating the effects of foreign innovation on the periodical. Furthermore, the sections show how *Scientific American* represents the changes in technology during the period. Finally,

they discuss the importance of viewing *Scientific American* and how research involved in *Scientific American* will allow scholars to better understand the general history of technology.

5.4.1   European Technology

During America's period of rapid industrialization in the mid-nineteenth century, industrial innovation in England and France was continuing to occur; however, at a more moderate pace than parallel development in America. *Scientific American* reflects European innovation during this time period through various articles related to European scientific achievement. Volume 2 of *Scientific American* mentions 1,203 European discoveries (*Scientific American*). It was observed that most of the foreign scientists in *Scientific American* were researchers rather than inventors and that most of the inventions in *Scientific American* were attributed to Americans while most research was attributed to European scientists.

One example of European achievement mentioned in Volume 2 of *Scientific American* is Professor Faber's speaking machine or Euphonia. This machine is designed to have the appearance of a human. It is capable of producing sounds similar to human speech, singing, or laughter. It was constructed after Professor Faber performed 25 years of extensive research on human speech patterns.

Another field of European research featured in *Scientific American* was astronomy. Many discoveries regarding astral body motion were attributed to European researchers. Articles within *Scientific American* also refer to European observatories and astronomical lecture series. One astronomical researcher mentioned in Volume 2 of *Scientific American* was William Parsons, the third Earle of Rosse. Parsons is known for constructing the world's largest telescope in 1847, which remained the largest telescope for the rest of the century. Using this telescope he was able to carry out pioneering astronomical studies that involved cataloging a large number of galaxies and discovering the spiral nature of nebulas.

A further area of European research mentioned in *Scientific American* includes the work of Sir George Cayley who today is considered the 'Father of Aerodynamics.' Cayley is an English scientist who constructed some of the first manned gliders as early as 1799. His research on projectiles is mentioned in Volume 2 of *Scientific American* along with accompanying illustrations that show diagrams of his projectile designs.

Although European research is more commonly mentioned in *Scientific American* than technological development, some European inventions are revealed. In general, these inventions are centered on the fields of transportation and agriculture. In transportation, the focus was on railroad travel, while the agricultural focus was on farming equipment. One of the more memorable technological innovations was the Atmospheric Railway which connected Dublin to Kingston. It is powered by air pumps rather than steam engines.

When European inventions were mentioned, the editor usually expected American inventors to quickly take and improve the idea. The editors often stated that Americans could produce better implementations of the European inventions due to the greater American need for the railway and farming technology. The editors often hinted that simple improvements could be made to the European inventions in order to make them more suitable to American needs as well.

Overall, in addition to reporting American innovation, *Scientific American* also gives a good overview of European scientific development during the nineteenth century. By the mid-nineteenth century, although technological innovation continued to occur in Europe, it generally was not suited to the American environment. Inventions that were made in Europe often needed alterations to be used by Americans. For this reason, the editors of *Scientific American* seem to focus more on the high-level research that was occurring in Europe during this time period.

5.4.2 Technological Change Observed in *Scientific American*

During the mid-nineteenth century, a large power transition was occurring in America. The old methods of generating power such as horses and water wheels were rapidly being replaced with steam power. In the early nineteenth century, steam was still a relatively new source of power. In 1770, James Watt constructed his famous Watt engine, which made steam engines more efficient. By 1780, the Watt engine had been adapted to textile mills. During the early nineteenth century, steam power became a reliable, safe, and efficient source of energy; however, it took a further half-century to convince mill owners that converting to steam would be beneficial. In the early volumes of *Scientific American*, it is difficult to see the progress in the development of steam and electricity or the decline in use of water wheels; however, it is possible to view

changes in regard to horse power.

An important source of power during the early nineteenth century was horses. Horse power machines, also known as horse mills, harnessed the strength of horses into rotational motion. This motion was then used to supply power to machinery that had previously been powered by hand. Horse mills had been popular sources of power since medieval times in Europe. Despite their long history of use, the nineteenth century continued to improve on the construction of horse mills to make them more efficient and easier to use. Many horse mill designs in the early volumes of *Scientific American* aim to increase the efficiency of the mill by having the center piece of the horse mill continue rotating, and therefore continue providing power, even once the horses have ceased motion.

By the mid-nineteenth century, horse mills were rapidly being replaced with steam power. In general, steam engines were easier to accommodate than horse mills. They did not require managing animals, and could be run nonstop without the fear of harming horses. The conversion from horsepower to steam power can be clearly seen in the early volumes of *Scientific American*, which were published at the end of the transitional period between these two sources of power.

In the early issues of Volume 2 of *Scientific American* (published in 1846), several improvements in horse power machines are described. An example is Woodbury's Horse Power, which can be found on the cover page of Issue 23 or Clark's Fraction Horse Power, which is located in Issue 44 of Volume 2. By the end of the volume, horsepower machines are still described; however, it is more common to find articles based on steam powered devices. For example, there were 3 horse power machines described in detail in Volume 2 of *Scientific American*. In this same volume, there were 9 steam engines described with illustrations. All the articles on steam engines can be found in issues following Issue 24 of *Scientific American* (*Scientific American* Vol.2).

The editors of *Scientific American* also comment frequently by the end of Volume 2 (published in 1847) that machines should be designed to be powered by steam sources rather than continuing to depend on horse power machines. In Issue 49 of Volume 2, Guyon's Lever Press is described by the editors of *Scientific American*: "Guyon's press used for pressing cotton. It is intended to be operated,-driven by horsepower, such being the way in which it is now generally used, but it is soon to be worked by steam, and this important power will undoubtedly render it more valuable" (*Scientific American* Vol.2,No.49). By Volume 6 (published in 1850), horse power machines are no longer mentioned in articles of *Scientific American*, which demonstrates their rapid replacement

with steam power by the mid-nineteenth century (*Scientific American*).

5.4.3 *Scientific American* and Technological Studies

Examining volumes of *Scientific American* from the mid-nineteenth century emphasizes the difference between technology considered important in the nineteenth century and the technology considered important today. For example, during the mid-nineteenth century, the possible uses of electricity were still unknown to many individuals. Electricity was seen as a scientific curiosity rather than an essential tool for modern life. For this reason, electrical achievements, such as Siemen's electrical telegraph developed in 1847, were not always mentioned in *Scientific American* upon their development. Volumes 2 through 9 of *Scientific American* include major series on common mechanical phenomena, but the corresponding series on electrical facts were much shorter with limited topics. An example of such a mechanical series is Mechanical Movements, which is present in Volumes 2 and 3 of *Scientific American*. A corresponding electrical series is Electrotyping and Electro-gliding which lasted for only 5 issues. Electricity was not a major part of the time period despite the suggestions to explore this field.

Many articles in *Scientific American* explain technology that was important to nineteenth century individuals, but is no longer important in modern day life. One common example of important technology for nineteenth-century individuals was a blind adjuster or sash fastener, which is a device that makes opening and closing window blinds easier. In current day, windows have been redesigned so that the nineteenth-century design for adjusters is unnecessary.

Another invention that was important to nineteenth-century individuals that is no longer studied includes perpetual motion devices. In modern day, it is understood that mechanical devices cannot run eternally due to the laws of thermodynamics. In the early volumes of *Scientific American*, articles regarding perpetual motion devices are common. The history of the development of these devices can be seen in *Scientific American* due to their importance and potential benefit to nineteenth-century individuals.

In general, through studying the articles of *Scientific American*, it is possible to observe inventions that were lost in history. For example, in the nineteenth-century, wooden carriages were still an important method of travel. Cars did not exist and trains were not practical for short traveling distances. Due to the importance of carriages, carriage wheel and axle designs were a common theme

of many articles and illustrations in *Scientific American.*

A final area observed was farming equipment that could be pulled by horses or pushed by men. Many simple devices were described in articles of *Scientific American* that would allow a farmer to perform the whole planting sequence by himself. This includes digging furrows, sowing seeds, and gently covering the seeds when the process is complete. Other machines were designed for simple harvesting. Although this technology was important to individuals of the mid-nineteenth century, it became unimportant in later years as steam-powered farm equipment was created and rendered the simple multitasking tools of the nineteenth-century unnecessary.

In general, through comparing *Scientific American* with the history of technology, it is possible to contrast commonly studied technology with the technology that was important to nineteenth-century individuals. Through this process, it is possible to better understand the history of technology by learning about the technology that was actually important during the nineteenth century.

Chapter 6

Future Work

Although the work conducted during this project has created a solid base for an image database, there are many possibilities for future expansion. It was understood at the beginning of the project that a complete image database of all nineteenth-century American illustrations would be impossible to achieve. For this reason, the database and cataloging scheme was specifically designed to allow the database to be expanded in the future.

As of May 2008, the image database contains eight volumes of *Scientific American*. It is hoped that future groups will expand this to include more volumes of *Scientific American* as well as multiple periodicals. *Scientific American* is primarily a mechanical periodical that describes patents that were being filed in the patent office in the nineteenth century. It is hoped that other scientific periodicals that include other aspects of science will also be cataloged in order to add diversity to the database. Furthermore, it is hoped that future groups will conduct research in order to try to find and catalog the first volume of *Scientific American* as well as other missing periodicals.

At the beginning of the project, the group conducted experiments that involved displaying an image of the illustration along with the corresponding metadata. In other words, an image search would result in a description of the image and also the actual illustration. Because of the amount of effort simply cataloging took, there was not enough time to attempt to extract all the images from the periodicals. There was also some question about Cornell's potential copyright on the digital copy they produced of the publications. All group members believed that the images made the database more interesting for a general user and would also aid a potential researcher; therefore, we think

that future groups should attempt to add this feature. From the initial research conducted into this aspect of the project, it was found that it took between 2 to 3 hours to extract the illustrations from the issues of a single volume of *Scientific American.*

The final area that was experimented with but not fully used in the image database was the idea of multiple subjects. We realized during the cataloging stage that some illustrations have multiple subjects that describe them (this is discussed in detail in Section 4.3.3). Because of differences in priorities dealing with cataloging, only two volumes were cataloged using the technique of multiple subjects. In the future, cataloging using multiple subjects should be applied to all the images in the database in order to make it easier to find images.

Further research should also be conducted in the future to better understand the periodicals cataloged as well as the history of printing and publishing. This project began to study these topics; however, for a complete understanding of early American illustrations, these topics should be expanded on. On the image database website there is space designated for the history of the periodicals. Once complete research has been conducted on this topic, it will be possible to post the results of the research on the website, helping database users better understand the images as well as the purpose of the particular periodical they appeared in.

The research that our group has conducted into the history of technology during the nineteenth century has given us insight into improvements that could be made to the properties in the database. While we had decided including large amounts of information on inventors and inventions was out of scope, an expansion of the database to include these is an option. When conducting research on foreign inventors, it was realized that the nationality or origin of the inventors or inventions may be beneficial to include in the database. Furthermore, the type of the invention (physical versus theoretical or conceptual) may also be beneficial for a researcher interested in the history of technology rather than the history of illustrations.

When the project began, a central question was the amount of time needed to catalog one volume of a scientific periodical. We assumed that a volume could be cataloged in a two week period as long as a few issues of the volume were cataloged each day. We found that the process of cataloging images takes longer than expected. The volumes of *Scientific American* that were cataloged in this project took about 50 to 55 hours to catalog completely. Revisions must then be made to check for errors that may have arose in the initial cataloging.

This process takes between 4 to 5 hours to complete properly.

The minimum amount of time that should be allotted to cataloging an individual issue of *Scientific American* is 20 minutes. In order to properly catalog an image, the article regarding the image should be read and understood. Even if the complete article is not read, it is often necessary to read a good portion of the article to truly understand the illustration that is being described. Some articles are written better than others; therefore, some images are easy to describe while other illustrations could take hours.

A suggested method for making the cataloging stage less time consuming would be to create an online interface to the database. This interface should be designed to allow individuals cataloging periodicals to enter cataloged data directly into the database. This would eliminate the intermediate step used in this project, which involved data being saved in Excel or text files before being inserted into the database. Eliminating a middle step will also save time in cataloging and managing the database.

A final suggestion for those interested in continuing the project in the future is to examine the Catalog of American Engravings available on the website of the American Antiquarian Society. The functionality of this catalog of images published before 1820 is similar to the functionality that our image database should eventually have. While our database is close to providing services similar to this database, there are still areas that could be improved such as adapting the user interface of our database to be similar to the American Engraving online catalog.

Chapter 7

Conclusions

At the completion of cataloging and conducting research, our group compiled our findings in order to provide direction and insight for future work on the project. Conclusions were made in regard to the database such as the database structure as well as the amount of time that is necessary to catalog a periodical. Other conclusions regarding the database discuss issues of data collection. Further conclusions were made into the research performed during the project.

After cataloging eight volumes of *Scientific American* it was possible to assess the amount of time necessary to catalog one volume of a periodical. Each issue of *Scientific American* was 7 pages long and contained between 6 and 9 illustrations spread over 3 pages. We found that the number of illustrations in each issue increased over the volumes of *Scientific American* studied. In other words, the early volumes contained between 3 to 6 illustrations while the later volumes contained more illustrations with greater detail.

At times several illustrations in an issue were dedicated to the same topic. In general, issues that contained many illustrations regarding two or three topics could be cataloged in 20 to 30 minutes. Issues that contained several illustrations regarding four or more topics could take 30 minutes to an hour to catalog. Some illustrations are described in extreme detail in *Scientific American* while other illustrations have vague descriptions; therefore, it is sometimes necessary to conduct outside research on a particular illustration. This process could result in a single illustration taking 1 to 2 hours to catalog. If very few cataloging problems occur, a complete volume could be cataloged in 55 to 60 hours. This figure includes the time needed for minor revisions. It was found that major revisions in content, which often require re-viewing a particular volume

took between 12 and 24 hours to complete (depending on the revision that was required).

In order to better understand the general work completed throughout this project, research was conducted regarding *Scientific American*, cataloging, and the history of technology in the nineteenth century. The research performed in regard to *Scientific American* revealed that *Scientific American* was a periodical that reflected the time period in which it was published. It was designed to discuss how events in the world affected theories and inventions being produced. It was further observed that the periodical evolved as time progressed to meet the needs and wants of its readers.

Through the research completed on the technical work done in this field, current work on knowledge representation was found to be relevant to projects such as this one. The Resource Description Framework in particular provided a preexisting and highly extensible abstract and concrete basis for this and future projects. Future groups should attempt to gain an in-depth understanding of the work done in metadata handling, and draw on the experience and work of communities such as the Dublin Core Metadata Initiative and the Semantic Web activity of the W3C, before beginning the actual work of a cataloging activity.

Also, while this project gained some of its separation from the standard variety of historical database by focusing relatively exclusively on the illustrations in *Scientific American*, a substantial portion of the work came to be dedicated to compensating for the implications of excluding a catalog of articles as out-of-scope; indeed, the project could have required substantially less work in constructing descriptions of the subjects of illustrations if this information was contained in metadata about articles, and in an included copy of the text of the articles themselves. Issues with linguistic shifts and searches can be substantially abated by the use of synonym lists to rewrite or expand queries to the vocabulary used at the time of writing. Full-text articles also offer the opportunity to produce a proper digital copy of the publication, with embedded metadata and text represented with text, which would be more conveniently accessible for future investigation into the period. It is recommended that future projects consider their purpose to be providing metadata for, and therefore improve the accessibility of, all of the information contained and directly related to their scope, and not try to catalog only one format in isolation of a mixed-format resource like *Scientific American*.

By studying the history of technology, it was found that the images within *Scientific American* parallel the evolving technology of the mid-nineteenth cen-

tury. This research further allowed the group to achieve a better understanding of the context of the images found in *Scientific American* and therefore gain a more in-depth knowledge of the illustrations cataloged. Studying the relation between nineteenth-century technology and *Scientific American* also resulted in a better understanding of nineteenth-century technology in general. This is due to the fact that observing the images of *Scientific American* allows a researcher to observe technology through the eyes of a nineteenth-century individual, which is a different perspective than is achieved from simply studying how technology progressed to its current state.

Overall, the database created and research conducted is only a beginning to this vast project. The results from both areas provide insight into how the project should proceed in the future. The next stage of the project should focus on expanding the current content of the image database through using the timing specifications and other results provided in this report. It will then be possible to conduct research and expand upon the conclusions found in this project. All in all, it is expected that through future project work, the image database designed and partially filled in this project will be expanded to become an essential tool for those studying nineteenth-century illustrations.

Appendix A

Database design for scientific illustrations

A.1 Goals

The overall goal of the database is to facilitate people finding the illustrations they want in scientific journals; to do this, some structure was considered desirable.

It was also a goal that the database be relatively straightforward to mutate and to add new information to, so that the work will be relevant to more than what can be cataloged over the course of three terms.

A.2 Techniques

At first, we used an SQL database to store our data. To design this database relatively standard Entity-Relation modeling was used; a set of appropriate and relevant entities was identified, the relevant relationships between them were identified, and a model drawn up. That model is reproduced in figure A.1.

The next step was to produce a definition for an SQL database from this model, which mostly involves assigning data types and putting foreign keys where they belong; for relations that are one-to-many, this means that the key for the one side goes in each row on the many side, one-to-one relations might as well be represented by both entities being represented in the same table, and many-to-many requires a separate table for the relationship.
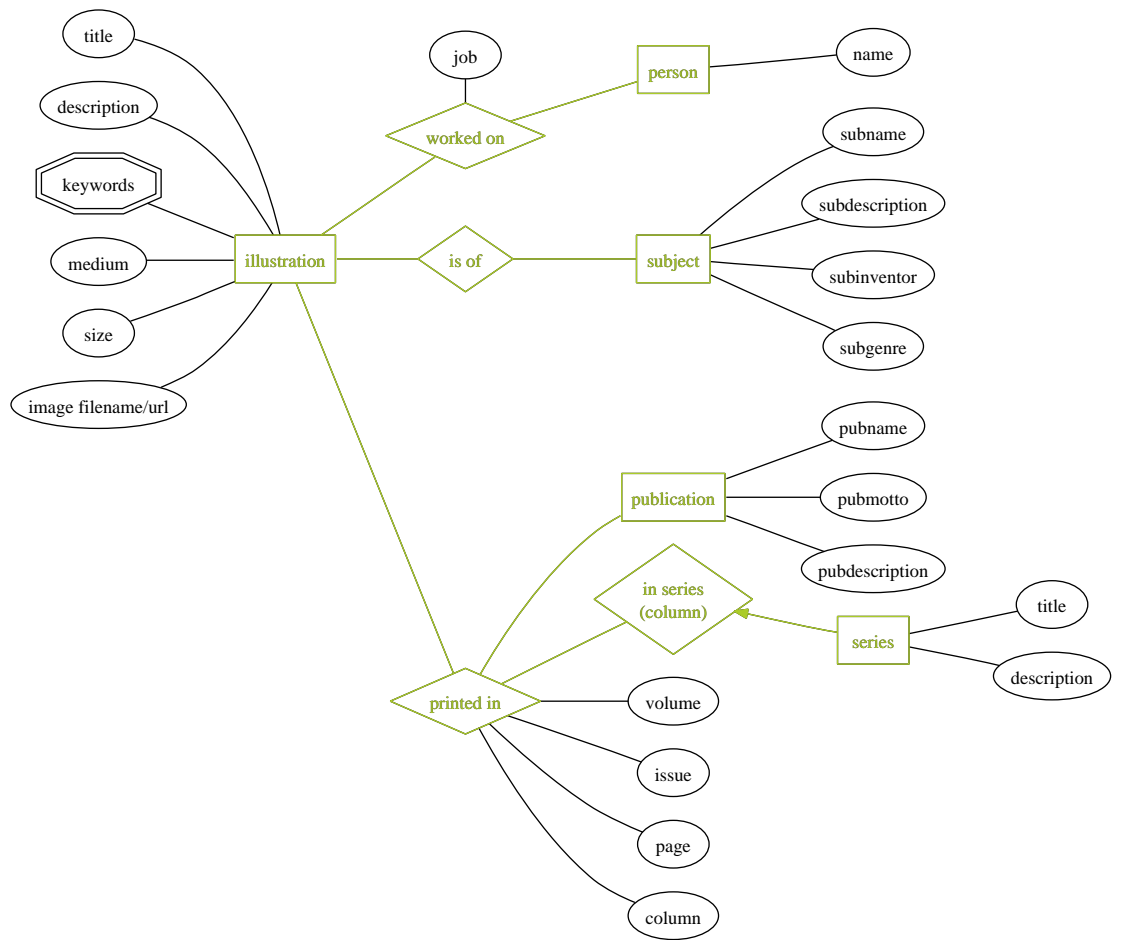
Figure A.1: ER model of database

A.3 Results

The ER model presented in figure A.1 is probably the best representation of the SQL database for a layperson; an entity-by-entity and relation-by-relation description seems appropriate. Straightforward entities are not included in this description.

This ER model also represents the RDF store used at the end of the project, with the modification that there can be any number of every attribute of all entities. The entities remained the same between the SQL database and the RDF store.

A.3.1 Illustration

The illustration entity is really what we're all about here; we want people to be able to find illustrations relating to their subject of interest. What is perhaps more relevant to the design than what is here is what is *not* here; while the title of the image, a description, keywords, and medium are here, along with ideally some way of getting a digital copy of the image, but not present are things like where the illustration was published, who did work on it, etc, as these are about other entities or relations between entities.

This also extends to meaning that the description of the image does not actually need to include a description of whatever the image is actually of; only information specific to this image is needed - so a cutaway view of Mr. Smith's railroad axle should mention that it is a cutaway view of Mr. Smith's Axle, perhaps showing some particular feature of that axle, but should not contain a full description of the nature of that axle.

A.3.2 Subject

The subject entity is a thing that is represented in an illustration, be it a new design for a windmill or a technique for metal casting. This is one place where the structure of the database shows it's merit; illustrations may be of multiple subjects, and subjects can be present in multiple illustrations. An illustration of Mr. John Smith's new design for a waterwheel should have at least two subjects; one for this particular sort of waterwheel, and one for waterwheels in general. These are especially useful for establishing interconnections between related illustrations.

The information we are keeping about a subject is a title, a description, a creator, and a genre; the description gives information of what the subject is; for instance, for Mr. Smith's Axle, there is some description of the workings and nature of that axle here. This description should not be mentioning an illustration, as the subject is likely to be used for many illustrations; a good heuristic is that if you'd mention it with the thing set in front of you, then it's appropriate for the subject description.

A.3.3 Series

Series entities represent series of articles, like a column in a newspaper. An example would be the series on mechanical movements in two volumes of *Scientific American.* This serves both as a place to describe relevant series, and to link together illustrations in articles of the same series.

What we're tracking for this are just a title and description.

A.3.4 Worked On

Worked on relates people to illustrations; if, for instance, Charles did something on a given illustration (And as he is quite prolific, he well may have), there is a record in this relation indicating this. There is also an additional piece of information associated with this relation, which is what job the person did; whether it was engraver, artist, etc - which of course can be left blank if unknown.

A.3.5 Printed In

'Printed in', on the other hand, relates illustration to publications, and as such contains a great deal of additional information; this is where all of the where and when of the thing goes - the volume, issue, date, page, and column.

A.3.6 In Series

'In series' relates a given printing to a series; the series in the sense we mean is really more of a property of the article, and so of the particular publishing of an illustration, than of the illustration itself. In both stages this was part of 'printed in', as the event of being part of a series was a property of the publication of the illustration.

Bibliography

Barnhill, Georgina Brady. "About this Cataloging ¿ Introduction." 2008.
⟨`http://catalog.mwa.org:7108/help/caepintro.htm`⟩

Bechhofer, Sean, van Harmelen, Frank, Hendler, Jim, Horrocks, Ian, McGuinness, Deborah L., Patel-Schneider, Peter F., and Stein, Lynn Andrea. "OWL Web Ontology Language Reference." *World Wide Web Consortium* (2004).
⟨`http://www.w3.org/TR/owl-ref/`⟩

Berners-Lee, Tim, ed. "Notation3 (N3) A readable RDF syntax." (2008).
⟨`http://www.w3.org/DesignIssues/Notation3`⟩

Brickley, Dan, Guha, R.V. ed., and McBride, Brian, ed. "RDF Vocabulary Description Language 1.0: RDF Schema." (2004).
⟨`http://www.w3.org/TR/rdf-schema/`⟩

Cowan, Ruth Schwarrtz. *A Social History of American Technology*. New York: Oxford University Press, Inc, 1997.

"Dublin Core Metadata Element Set." *Dublin Core Metadata Initiative* (2008).

"DCMI Metadata Terms." *Dublin Core Metadata Initiative* (2008).
⟨`http://dublincore.org/documents/dcmi-terms/`⟩

Scientific American, vol. 1. New York, 1845.
⟨`http://en.wikisource.org/wiki/Scientific_American/Volume_1/Issue_1/Front_page`⟩

Scientific American, vol. 2-9. New York: Munn & Company, 1846-1854.
⟨`http://cdl.library.cornell.edu/moa/`⟩

"Outline of U.S. History, Chapter 6: Sectional Conflict." *International Information Programs USINFO.STATE.GOV* (2008).
⟨`http://usinfo.state.gov/products/pubs/histryotln/conflict.htm`⟩

Klyne, Graham, ed., Carroll, Jeremy J. ed, and McBride, Brian, ed. "Resource Description Framework (RDF): Concepts and Abstract Syntax." *World Wide Web Consortium* (2004).
⟨`http://www.w3.org/TR/rdf-concepts/`⟩

Kronick, David A. *A History of Scientific and Technical Periodicals (The Origins and Development of the Scientific and Technical Press 1665-1790)*, chap. 14. The Scarecrow Press Inc., 1976, 2nd ed.

Library of Congress. "About the Prints and Photographs Catalog (Prints and Photographs reading Room)." 2008.
⟨`http://www.loc.gov/rr/print/catalogabt.html`⟩

———. "Frequently Asked Questions (Library of Congree Online Catalog)." 2008.
⟨`http://catalog.loc.gov/help/faq.htm`⟩

McCorison, Marcus A. and Hench, John B. "A Brief History of the American Antiquarian Society." *American Antiquarian Society* (2007).
⟨`http://www.americanantiquarian.org/briefhistory.htm`⟩

Mott, Frank Luther. *A History of American Magazines*, vol. 2, chap. 8. The Belknap Press of Harvard University Press, 1967.

———. *A History of American Magazines*, vol. 1. The Belknap Press of Harvard University Press, 1970.

Prud'hommeaux, Eric, ed. and Seaborne, Andy, ed. "SPARQL Query Language for RDF." *World Wide Web Consortium* (2008).
⟨`http://www.w3.org/TR/rdf-sparql-query/`⟩

Schafer, Elizabeth D. "Scientific American." *St. James Encyclopedia of Pop Culture* (2002).
⟨`http://findarticles.com/p/articles/mi_g1epc/is_tov/ai_2419101070`⟩

Scientific American, Inc. "Scientific American: About Us: History." *Scientific American* (2008).
⟨`http://www.sciam.com/page.cfm?section=aboutus`⟩

Steinberg, S.H. *Five Hundred Years of Printing*. Baltimore: Penguin Books, Ltd, 1955.

The Schoyen Collection. "Pre-Gutenberg Printing." 2008.
⟨`http://www.schoyencollection.com/Pre-Gutenberg.htm`⟩

Thomas, Isaiah. *The History of Printing in America*. Albany, NY: Joe Munsell, printer, 1874, 2nd ed.