TELLING A DIFFERENT STORY:

THE HISTORY OF SCIENCE AS A WEB OF RELATIONSHIPS

IN THE EMBRYO PROJECT

by

Cera Lawrence

A Thesis Presented in Partial Fulfillment

of the Requirement for the Biology and Society Capstone Project

ARIZONA STATE UNIVERSITY

May 2009

ABSTRACT

The Embryo Project is a history of science project that uses a data repository architecture named Fedora to create an online collaborative encyclopedia. Only one of a number of projects using this new open source architecture, the Embryo Project seeks to exploit Fedora support of semantic relationships to facilitate collaboration among researchers, to encourage deeper and more extensive reading by the non-expert public, and to reveal unexplored connections in the history of embryology between people and social contexts. Does it work? Does developing an online, object-oriented, semantic-relationship-rich data repository for the history of embryology give us something that more traditional history projects do not? Or is it just a website? This paper analyzes the utility of Fedora from the perspective of an undergraduate researcher, and demonstrates that there is more to the investigational strategy of interconnected relationships than content generation for a web encyclopedia.

ACKNOWLEDGEMENTS

My mentor and first reader, Jane Maienschein, provided invaluable assistance and encouragement. Grant Yamashita, my second reader, helped make sense of the technical aspects of Fedora and is still assisting in finding good sources for the discussion of semantic web technology. My third reader, Mary Sunderland, asked the key initial questions that framed this project. Marie Glitz, editor of the Embryo Project Encyclopedia, made herself available for numerous pep talks.

This research was supported in part by funds from the Howard Hughes Medical Institute through the Undergraduate Science Education Program, by the National Science Foundation through the Center for Biology and Society, and from the ASU School of Life Sciences.

Table of Contents

1. Introduction and Background

The Embryo Project is a collaborative effort to create an online repository and encyclopedia of information relevant to the history of embryology. In partnership with the Arizona State University Library, the Embryo Project uses software called Fedora[1] (short for "Flexible Extensible Digital Object Repository Architecture") to store and retrieve content. The Fedora software, used by the University Library for a number of repository projects, was specifically designed to take advantage of an emerging information organization technology called semantic web. The semantic web is composed of data objects and their semantic relationships with each other (Herman, et al. 2009). Semantic relationships are a way to encode more real world information in a database by describing how one particular thing is related to another particular thing (Storey 1993, Allemang and Hendler 2008). A semantic web is the web of data created by these relationships and the things that they connect. Workers in the information management field often refer to "the Semantic Web" in the definite article, reflecting the theoretical prediction that all data everywhere could be connected to one web (Herman, et al. 2009).

Semantic web technology seems to hold a great deal of promise for research in the history of science. Most narratives of the history of science follow a linear structure, spotlighting the researchers, experiments, and events that have already been well established to be important (Pinto-Correia 1998, Griesemer 2007). But this technology for storing and sharing information offers a new method for the process of historical research (Lagoze 2006). As a result of the use of semantic web technology, the

[1] Fedora, the object oriented database software should not be confused with Fedora, a version of Linux distributed by Red Hat. Due to trademark concerns, Fedora has been rebranded as "FedoraCommons." This paper refers to "Fedora" throughout because it is less awkward.

researchers (including undergraduate trainees) who produce the entry level content for the online Embryo Project Encyclopedia ("the Encyclopedia") are trained to find information about the relationships among people, places, technology, experiments, and concepts.

I became interested in the Embryo Project when I attended a social gathering at the Center for Biology and Society at Arizona State University. Dr. Manfred Laubichler took the opportunity to show a poster about the Project to me, and I was immediately impressed by the idea that the encyclopedic entries would be linked to other entries--and not with hypertext links, but with meaningful, semantic links. The database and website would know not just that two things were linked, but the nature of the real-world link between the subject of one entry and the subject of another entry.

This aspect of the project impressed me because it was similar to a project I had worked on in a previous corporate job. I worked for several years as the Information Management Director at a professional membership organization called the International Society of Six Sigma Professionals[2] (ISSSP). One of our membership products was called a Professional Profile, and this online site was meant to catalogue all the professional recognition each member had received, the history of the member's interaction with the Society, and the member's interaction with other members of the Society, both professional and corporate. One of the ways in which we had sought to create this Profile was to encode each member's "relationships" with companies at which they had been employed and events at which they had spoken, rather than making each activity a simple line item on a web page. This would enable ISSSP's workers to quickly

[2] http://www.isssp.com

change the look of the Profile as necessary, and eventually sort and analyze information about trends of product usage, event attendance, and project management.

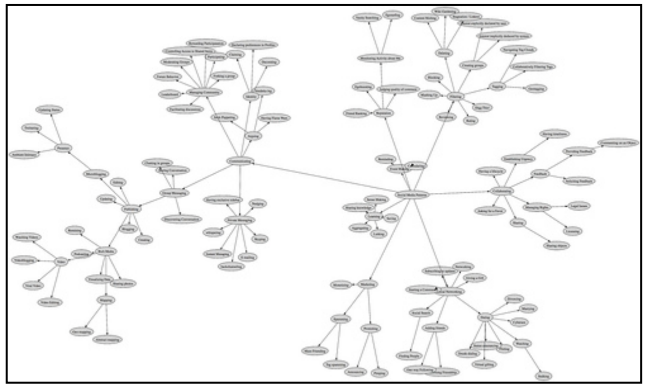
The ISSSP project never quite achieved its goals, but it always felt like a worthy endeavor to me, and an interesting way of managing content. I was, therefore, excited by the prospect of working on a research project that was using the same theoretical basis as the Professional Profile, but was engaged in exploring the relationships between actors in the history of science. It was immediately clear how interesting the view of history as a network of influences and interactions could be. I thought of the utility of dynamically created diagrams outlining the flow of scientific developments in the field of embryology, and was excited to facilitate the production of an online encyclopedia that contained more information than just the single layer of reference material.

As I have worked on the project, I have found it difficult to explain to others just what it is that the Embryo Project does. When I say that it is creating an online encyclopedia of the history of developmental biology, they immediately think of Wikipedia. The comparison is not a bad one. The Embryo Project peer review and publication process lend its Encyclopedia a scholarly weight lacking in Wikipedia, but the general principle of a publicly accessible online resource is the same. And yet, the model of online data repository that is the wiki system does not include the richness of fully defined semantic relationships as the links between each entry.

Imagine taking a well-linked Wikipedia article and zooming out from it, with each hyperlink becoming a spoke in a diagram that connects to other articles as nodes. From that "zoomed out" diagram view, you can see not only the primary links to other articles from your primary article, but secondary and tertiary links as well. Now imagine

Figure 1.1 A suggestion of what such a diagram might look like. http://www.flickr.com/photos/kentbye/

that each of those spokes in the diagram, each of those links, is not just a representation of a connection but also contains the information of what kind of connection exists. You can ask this diagram to only show



coworker relationships, or to only show relationships between persons and organizations, or between persons and persons, or to only show mentorship and student relationships. You can produce various views of timelines, or expand thumbnail images that are relevant to your search. Because of the extra information contained in the semantic web, your ability to work with all of the information contained in this encyclopedia is much improved. That is part of the goal of the Embryo Project Encyclopedia—improving the ability of both the end user and the Project researchers to work with information about the history of embryology.

This paper will discuss in detail how the Encyclopedia is put together and how semantic relationships are used in it, including the technical underpinnings of the Fedora database architecture. After making it clear exactly what a semantic relationship is and how the project works, I will delve into some of the actual results of performing historical research in this way. The ultimate goal of the Embryo Project, as outlined in the 2006 Proposal to the National Science Foundation Human and Social Dynamics Initiative, is to identify “Change Agents” in the history of embryological science: " those factors and forces that have affected and changed the scientific study of embryos." Through the use of semantic web technology in Fedora, the Embryo Project records information about those agents in history and their connections with each other. Using technology that helps dig into these relationships between agents is a novel approach to studying the history of science, and it seems that it ought to help deliver a new understanding of the way scientists work as part of society.

The main question that I seek to answer in this paper is whether or not developing this object oriented web of semantic relationships does, in fact, give the historian of science something that a traditional review of primary and secondary literature does not. Does the use of semantic relationships serve the goal of the Embryo Project as stated in the 2006 Grant Proposal? Does it generate new knowledge about the history of embryology? I am interested in the way we tell stories about the history of scientific endeavors, and how new ways of seeing those stories changes our approach.

2. History and Purpose of the Embryo Project

The Embryo Project began in 2006 as a joint project by Manfred Laubichler and Jane Maienschein, faculty members of the Human Dimensions in Science group in the Arizona State University School of Life Sciences. The intent, as a reading of the grant proposal to the National Science Foundation reveals, was to identify the "Agents of Change" in the history of embryology, the scientific study of embryos. The Agents of Change would be identified gradually as the Embryo Project produced information regarding how scientists have gone about conducting research on embryos, and the scientific and societal contexts in which their research took place. Identifying the agents that drive scientific research would be interesting in any scientific field of study, but the researchers chose embryology in part because of some of the clearly unsettled tangle of ethical and political controversies. As stated in the grant proposal: “embryo research is always embedded in webs of unsettled ethical, legal, political, religious, cultural, economic, and social negotiations that shape the conduct of science.” (Maienschein and Laubichler, "EP Proposal" 2006 p B-2).

The Embryo Project was conceived with many smaller objectives that feed into this one big goal. These objectives of include making the Encyclopedia content available to the general public, educating undergraduate students about history of science research, and providing a collaborative environment for multidisciplinary work. The objective with the most public impact is the Encyclopedia. As an online resource about the history of embryology, it is written with both the non-expert public and the academic community in mind. The Encyclopedia functions as a tool to educate the interested non-historian, or the secondary-school student, about the people and institutions involved in the study of embryos, from Aristotle to the present.

The second objective is to enlist undergraduate students as first-tier researchers to create the basic content for the Encyclopedia. With supervision from a graduate student and/or a post-doctoral fellow, selected students write entries for the Encyclopedia and perform other data collection and entry activities such as entering metadata for archival photographs and creating the spreadsheets that are used to produce timeline views. The decision to include undergraduate students in the content generation process has allowed the Embryo Project to have enormous educational impact at Arizona State University. Undergraduate students have created most of the content in the Encyclopedia to date, and many have been able to present the results of their work in the Embryo Project as original research, a graduation requirement for both the University Honors program and the

School of Life Sciences Biology and Society program.

The Project fosters a richly multidisciplinary collaboration with professionals in other academic fields and with other institutions. The multidisciplinary objective is necessary because the Principle Investigators intend to explore both the history of the highly technical aspects of embryo research and the history of the social movements surrounding and interacting with embryo research. People with expertise in ethics, policy and law, religious history, microbiology, zoology, and evolution (to name a few) are all needed in this endeavor. The Project's institutional partners are the MBLWHOI Library, and the Max Planck Institute for the History of Science.

3. Fedora as the Tool

The tool that enables the Embryo Project to accomplish these objectives is software named Fedora, which is short for Flexible Extensible Digital Object Repository Architecture. The Information Systems and Technology staff at the ASU Library, already using the software for other projects, pitched Fedora to the Embryo Project. Why was Fedora selected, and how does it help achieve the objectives discussed above? Part of the answer lies in the reasons Fedora was developed in the first place.

The World Wide Web in its original form displayed flat text files marked up with HTML (hypertext markup language), which were coded "by hand" and updated by replacing the entire content of a file with a new version. As the desire for dynamic content grew—web pages that were somewhat aware of changing information, and did not have to be updated manually—web developers began to hook their websites up to relational databases.

A relational database is one made up of several different kinds of data that relate to each other in various ways. There are three important terms used in describing how a relational database works: ‘relation’, ‘tuple’, and ‘primary key’. A relation is a set of data that are united by common attributes. For example, a database managing a library’s catalogue would probably have a relation named “Books” that would contain information about each book possessed by the library. That information (author, publishing date, and subject keywords) would be the attributes for “Books”. Each tuple, or row, in the relation contains the set of information about one distinct item in the group the relation refers to. The relational data model uses a primary key, which is the combination of columns (attributes) necessary to uniquely identify each tuple. This could be just one column, such as an id number, or multiple columns (Codd 1990).

Relational databases can contain multiple distinct relations. A common example is a retail database containing “Orders,” “Customers,” and “Inventory,”which interrelate. This prevents a lot of duplication of information that would occur if all that data were held in one table, as they would be in a flat text file. It is important to note that these interrelationships are not semantic in nature. They refer to data in other tables, but they do not specify the manner in which the data are related.

As websites began to make more sophisticated use of relational databases to present dynamic or user-tailored information, users began to expect multiple different kinds of content. First inline images, and then complex, interactive interfaces and embedded movie files found their way into the every day experience of a web user. The applicability of this sophisticated content for archival institutions like libraries and museums was obvious—online exhibits of digitally archived collections—but challenging to achieve. Relational databases were certainly capable of storing all the information needed to create these exhibits, but to display that information required a team of dedicated web developers. Fully developed so-called “turn key” systems, while inexpensive and easy to implement, usually were not flexible enough to meet a specific institution’s needs. Fedora was created to meet those needs (Lagoze 2006).

A joint team from Cornell University and the University of Virginia Library developed Fedora in its current form (Lagoze 2006). Fedora is an object-oriented database designed to store and organize digital content, and to do so in ways that are flexible and robust enough to meet the needs of a multitude of scholarly organizations. It is open source, which means there is no cost to use the software and anyone using the software may view the actual source code, which aids some trouble-shooting tasks. Fedora is also designed to use semantic relationships between different entities in the database, which adds an additional layer of information that a relational database is unable to store natively (Lubyte 2007).

Let’s explain briefly what these three traits—object-oriented, open source, and semantic web—mean in simple terms. Object-oriented is a term that comes from programming languages like Java and C that use modular code in order to avoid duplication of effort (Lagoze 2006). There are lots of good reasons to do this, which are covered at length in computer science textbooks. The basic gist of object-orientation is that there are classes of objects that describe a set of unifying characteristics, and then there are instances of each class that use and modify those basic characteristics in some way.

Though the particulars of object-oriented programming and its benefits are well beyond the scope of this paper, it should suffice to explain a little bit about what "objects" are in Fedora, and why an institution would want to use them in an online digital repository. An object is a discrete entry in the database. The component parts of an object are its datastreams and its persistent id (PID). The PID performs the same function as the primary key of a relational model, and is simply the unique identifier by which the system calls any object. The datastreams of an object are the digital encoding of the file or files that will be shown to the user when this object is called for. This is the full text of an article, the jpeg image of a photograph, the mpeg file of a movie, and so forth. Any object can have multiple datastreams as parts of that object, but every object in Fedora has at least a Dublin Core datastream, which contains its metadata in a standardized format (Lagoze 2006). Metadata are data about the object, usually used to make searching and organizing easier. For Embryo Project articles, this includes the title, author, search keywords, and posting date.

Because the system is object oriented, each basic object can be pulled together with other objects into composite views. For example, an article about Dr. Cornelia Clapp could be presented with several images of her working at the Marine Biological Laboratory, scans of her handwritten letters, and, were such a thing available, a short movie of her. These components are tied together with semantic relationships (Lagoze 2006), but remain distinct and separate and can be re-used as part of any other view without requiring any duplication.

Fedora is also an open source project. Free and open source software (FOSS) refers to software for which the developers make the source code openly available. Open source software developed as a response to the copyright and intellectual property practices by development companies such as Microsoft, Adobe, and others who keep the original source code of their products secret, and sell users a license to use a packaged and compiled version of the code. The model for open source software licensing is the General Public License, or GPL, which is a version of licensing copyright that entitles anyone to view the source code, compile the source code, and make changes to it. The GPL also provides permission for anyone to use any portion of the code, including the entirety, to make derivative software or entirely new software. The GPL only stipulates that the resulting derivative work must also be licensed under the GPL, so that the source code remains available to everyone (Lessig 2006).

Richard Stallman developed the GPL model when he developed his open source operating system, GNU, which together with Linus Torvald's Linux became the prime example of an open source software project, GNU/Linux. Stallman famously draws a distinction about the type of freedom represented by FOSS: "Free as in free speech, not free as in free beer." (Lessig 2006) It is not against the GPL to charge money for open source software, or to otherwise make FOSS a commercial venture. It is only a violation if the source code, in its entirety, is not published in such a fashion that anyone can compile and use the code if they know how to. This creates what some call an "open world.[3]"

The point is not to make it impossible to make money on software, but rather to open up the work to many different people to collaborate on and to adjust the software to

[3] This "open world" mentality strongly suits the Embryo Project's approach. The Embryo Project uses the Creative Commons (http://creativecommons.org) copyright license to manage its content, as do many other projects that use Fedora. Creative Commons is an organization founded to create strong, legal copyright agreements that specifically retain some rights to the content while explicitly granting others. This permits anyone to reprint in whole or part any part of the Encyclopedia's content, as long as they give proper attribution to the author and the publisher—and as long as they do not do so for the purpose of financial gain. Anyone is permitted to use the text of the Encyclopedia's entries, or the image files that accompany them, as long as they do not adapt, change, or build upon those works.

their own needs (Payette 2007). Fedora is open source software in order to increase the speed at which it can be tinkered with to suit certain projects' needs, and in order to take advantage of the myriad other ways in which open source improves a project.

Fedora's status as FOSS is useful for the Embryo Project because it means that our "in-house" programmers at the ASU Library can view every line of code. Source code awareness improves the feasibility of creating complementary add-ons or plugins that works with Fedora, and means that the programmers can, if necessary, change the code to suit the Embryo Project's needs (Lagoze 2006). But the characteristic of Fedora that is most important in achieving the objectives of the Embryo Project is its ability to store and manage semantic relationships, thus launching the Embryo Project into the information management frontier known as the Semantic Web.

Semantic Web and Semantic Relationships

Semantic relationships are an information management concept originally borrowed from philosophy. In this context, a semantic relationship is a data abstraction that encodes real-world information about one object's relationship to another object. (Herman, et al. 2009) It is close to "natural language," which is why it is termed semantic; any semantic relationship is essentially the verb phrase used to describe the relationship in speaking or writing about it. Thus, it is easier to model reality or store information about reality in the form of semantic relationships.

Semantic relationships are established in the Fedora system by asserting a relationship link from one object to the PID of the related object. In the Embryo Project, we assert the relationship within the text of an article—“`used`” for example—and that relationship is coded in indexable XHTML as

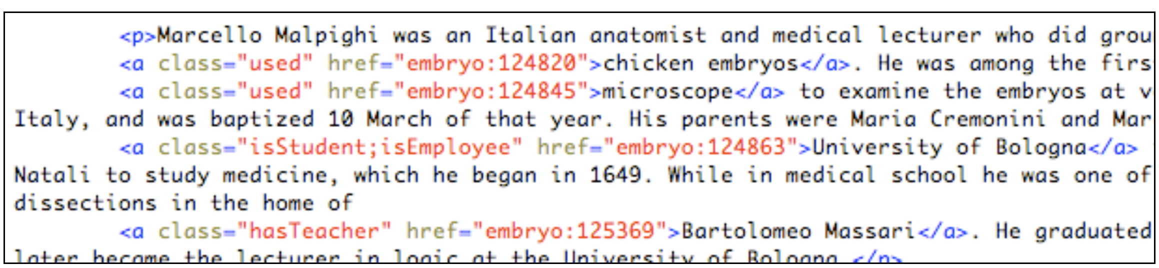
```
<a class=”used” href=”embryo:%pid%”>
```

where %pid% is the persistent id of the related object (See Figure 3.1 for an example).

Fedora grabs these assertions and stores them in a datastream called RELS-EXT, re-encoding them to the industry standard Resource Description Framework (RDF) syntax (Lagoze 2006; Herman, et al. 2009). A line-item entry in the RELS-EXT datastream might look like this:

```
<emb:used rdf:resource="info:fedora/embryo:124820"
rdfs:label="chicken embryos"/>
```

and refer to the fact that the subject of the object used chicken embryos in his or her research.



```
        <p>Marcello Malpighi was an Italian anatomist and medical lecturer who did grou
        <a class="used" href="embryo:124820">chicken embryos</a>. He was among the firs
        <a class="used" href="embryo:124845">microscope</a> to examine the embryos at v
Italy, and was baptized 10 March of that year. His parents were Maria Cremonini and Mar
        <a class="isStudent;isEmployee" href="embryo:124863">University of Bologna</a>
Natali to study medicine, which he began in 1649. While in medical school he was one of
dissections in the home of
        <a class="hasTeacher" href="embryo:125369">Bartolomeo Massari</a>. He graduated
later became the lecturer in logic at the University of Bologna.</p>
```

Figure 3.1: Example of asserting a relationship class in an anchor tag

The RELS-EXT datastream holds the relationship information so that it can be used to produce views of and information about the semantic web thus created. Using RDF to do this means that the information can easily be re-indexed into a special database called a triple store. The triple store is named such because it stores the three pieces that make up a single semantic relationship: Object A, the relationship verb phrase, and Object B (Fedora Commons 2007). Triple stores are useful because they permit easier handling of semantic relationships. This allows, for example, diagrams of semantic webs to be dynamically called into view. Each component object is also part of myriad other potential webs. The web that is spun out of the triple store depends on what characteristics are defined in the query. Views of the semantic information can be created from the different modular objects as needed, on demand.

This aspect of Fedora is what facilitates the Embryo Project's stated goal of identifying Change Agents in the course of history. Semantic relationships can reveal previously unrecognized information about the history of embryology because the system stores every connection that the researchers on the project collaboratively make. The diagrammatic views that can be created with semantic relationships permit exploration of the webs of relationships. Really robust services, currently in development, will be able to indicate where integral gaps in the Encyclopedia's content lie. But even though these services are not fully developed, the conceptual framework of the Embryo Project's semantic relationship ontology has already proven to facilitate research.

Semantic relationships in the embryo project

Any system using semantic web technology must use a domain-specific ontology[4], which is essentially the controlled vocabulary of the objects and the

[4] "Ontology" is a misleading term in information management. It is derived from the term as used in metaphysics, meaning 'the study of the nature of being, reality, and existence' or, as the Stanford Encyclopedia of Philosophy has it, simply "the study of what there is" (Hofweber 2004). Tom Gruber, of the Stanford University Computer Science Department, introduced the term to knowledge and information management in 1992, as a useful way to describe an agreed-upon, shared vocabulary to represent objects and their relationships (Gruber, Translation Approach 1993) (Gruber, Design of Ontologies 1993). The term has remained the source of some small amount of controversy in what is colloquially known as "the blogosphere," prompting such articles as "When "Knowledge Engineers" say "Ontology" They Mean The Opposite: "Epistemonomy". (Harnad 2008) Gruber states that while the term comes from metaphysics, the AI application is "certainly a different sense of the word than its use in philosophy" (Gruber, What is an

relationships that will be used in that system. (Herman, et al. 2009) Domain-specific means that it delineates the only kinds of relationships that can exist between the kinds of things the database contains. The ontology of a semantic web describes the manner in which objects' relationships are conceptualized (Allemang and Hendler 2008). The Embryo Project's ontology domain describes the ways in which the various categories of objects in the Encyclopedia can be related to each other. Its current incarnation is recorded in "The EP Ontology Sheet," provided in Appendix A.

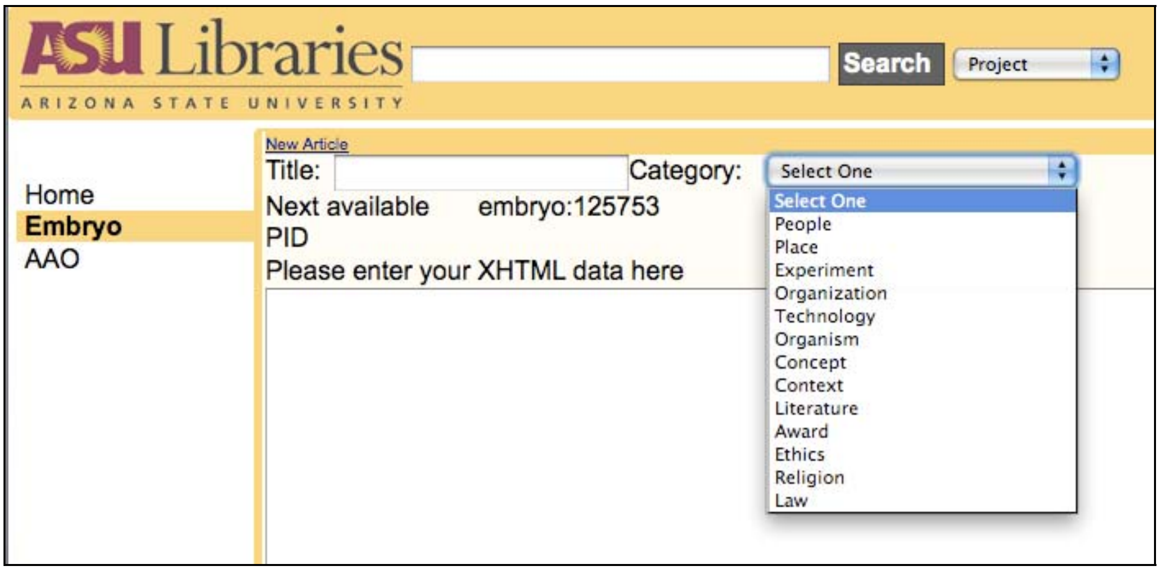


Figure 3.2: The Object Posting Application, and the list of categories

The encyclopedia entry authors in the Embryo Project must deal with semantic relationships in their articles first by determining the Category of the subject. The EP Ontology sheet (See Appendix A) lists our categories of articles: People, Places, Organizations, Concepts, Organisms, Experiment, Technology, Law, and Literature, to name the most commonly used. The category is asserted in the XHTML of the article by a meta tag:

Ontology n.d.). While the use of the word ontology does convey the important idea that semantic web technologies deal with abstractions of reality, the term can be supremely difficult to wrap one's head around.

```
<meta name=”dc.subject” content=”People”>
```

and is also assigned at the point of object creation in the posting web application (Figure 3.2).

The category of the entry subject being determined, the author must then determine the category of any object to which the entry links. This determines which relationship type can be used between the two objects. By referring to the Ontology Sheet, the author decides which relationship is appropriate for this link. The direction of the link is important, and a basic knowledge of sentence structure is helpful in this. A semantic relationship can be thought of as a short declarative sentence in English. The subject of the entry is the subject of the sentence, the relationship is the verb phrase, and the linked-to object is the object of the sentence.

In this way, the ontology is established by the real world constraints of how things really relate to other things. Many relationships ought only to be used between a subject of one type and an object of another. As an example, `Malpighi used chickens` is information that makes sense, but the relationship cannot go the other way, to say that `chickens used Malpighi`. More than just subject and predicate concerns are important, because only certain kinds of relationships make real world sense between certain types of objects. I may wish to say that Albrecht von Haller `hasStudent` Charles Bonnet, but I do not want to say Albrecht von Haller `obtainedDegree` Charles Bonnet, because the relationship ‘`obtainedDegree`’ does not apply between two people. This is an absurd example so that the problem is clear, but there are other examples in the ontology that require very fine distinctions. If the semantic web formed by relationships does not make real world sense, the value of the data entry is nil at best and at worst, wrong.

The use of relationships is one of the aspects of writing for the Embryo Project that has a learning curve for the undergraduate researcher. Undergraduate researchers enroll for course credit in a semester-long seminar during which they are trained in writing entries for the Encyclopedia. The training seminar often spends a lot of time explaining how semantic relationships work, and even experienced researchers use them incorrectly at times. However, the conceptual framework greatly benefits the researcher once it is clear, and promotes high quality articles and speedy content generation. The object categories focus the article writer's research. The researcher must conceptualize the article he or she is writing as only about a Person, or a Technology, or an Organization.

Focusing on just one entity at a time empowers the researcher to produce a complete treatment of that one subject without getting bogged down by extraneous details such as the complexities of a particular technology's public reception. However, not one of these details is lost to the project because the researcher can refer to them with relationships—"`technology contributedTo complex-social-context`". Article writers are encouraged to think about and log all of the relationships they find as they research their narrowly focused entry, and to create a cluster of eight closely related articles for their first independent project. Each article they write creates dozens of potential other articles, and completes the semantic relationship links from articles previously written.

The Embryo Project jargon for this approach is "atomistic"; Fedora's developers call it "Lego-like." Others call it "modular." Jargon notwithstanding, this approach is the strength of an object-oriented, semantic-web-aware, online history of science project. The end result is speedy content generation of focused entries that relate to multiple other entries in the project. Eventually the relationship sets can be queried to produce views of the related objects. The relationships also create a way for collaboration to take place asynchronously. The nodes of various webs may comprise entries written months or years apart from each other.

4. Case Study of the Embryo Project's Undergraduate Impact

The 2006 NSF Grant Proposal describes one of the intended impacts of the Embryo Project as the recruitment and training of undergraduate researchers. These undergraduates would work as directed by the PIs of the project, performing basic historical research to support the PIs' own scholarly projects. In the spring of 2008, I applied to and was accepted to the Embryo Project Seminar, where I began learning how to write entry-level articles for the Encyclopedia.

Undergraduates are encouraged to find a particular aspect of the vast history of embryology that they find interesting, and devise a cluster of eight articles that deal with that aspect. I was interested in the theory of development known as Preformationism from the late seventeenth through eighteenth centuries, since a lecture in my human sexuality class had mentioned ovism and spermism in passing. I was curious to know more, and Drs. Maienschein and Yamashita gave me the okay to develop a cluster of articles about Preformationism.

As many undergraduate researchers do when they first attempt to write for the Embryo Project, I started with very big ideas that were too unwieldy to do well and quickly. The first article I presented to be workshopped in the Seminar was a Concept-Type article entitled “Preformationism.” While the article was reasonably well written, it was clearly not a thorough treatment and suffered from being a little choppy in its transitions, as well as being far too long. My peer reviewers, other undergraduates taking the seminar, dispensed much good writing advice, and Dr. Yamashita suggested that rather than try to write such a big concept to begin with, I should write some biographical entries first. Dr. Maienschein agreed, saying something along the lines of, “This is too hard to start with. Chop away at a few related things and then come back and do an interpretative essay.” The connections I would find in the biographies, they suggested, would lead to other subjects to write about, and after completing a few entries pertinent to the development of Preformationism as a theory, I would be in better shape to successfully write the Concept article.

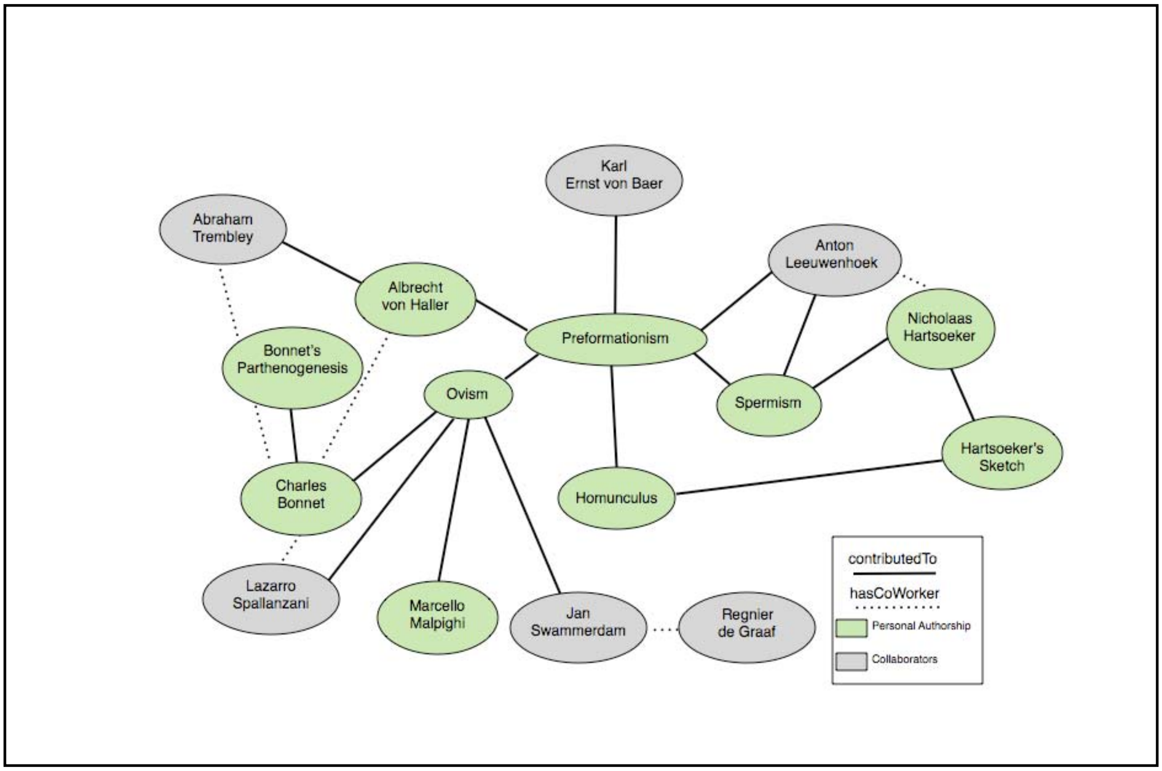


Figure 4.1: Representation of the semantic web built from the preformationism cluster. Personal articles are in green bubbles, and collaborators' articles are grey. There are more connections than are shown in this web, but this is a reasonable approximation. The semantic relationships, represented by the lines, are taken directly from the XML code for the articles.

I stepped back from the too-hard Preformationism article and worked instead on a biography of Nicholas Hartsoeker (a spermist), and then an article about the iconic sketch of a homunculus drawn by Hartsoeker in *Essai de Dioptrique*. The research I did for these two articles led to an article about the concept of the homunculus, and then to two more biographical articles: Marcello Malpighi and Albrecht von Haller. After writing these five articles and reviewing the entries that already existed in the Encyclopedia to which my five articles linked, I was ready to write the entries for ovism, spermism, and finally Preformationism. With plenty of background work done to support these big concept articles, my final products were informative, well-focused, and well integrated into the Encyclopedia's relationship/semantic web. (See Figure 4.1)

Over the summer, as an employee of the Embryo Project through the School of Life Sciences Undergraduate Research program (SoLUR), I wrote two more articles about the development of Preformationism: Charles Bonnet's biography, and an article about his experiment that proved parthenogenesis occurs in aphids. These two articles fit very nicely with my overall cluster, with many relationship links that tied them into the semantic web I was seeing take shape. Under Dr. Maienschein's direction, I undertook an entry about Aristotle's contribution to embryology, which would provide both a chronologically foundational entry for the Embryo Project Encyclopedia as well as background for the Preformationism cluster. This entry was originally to be a biographical piece focused on Aristotelian biology, but it quickly became apparent, based on my experience with the Preformationism cluster, that a biography about Aristotle would be difficult to write at the proper scope for the Embryo Project. Therefore, I chose to refocus the article as a Literature piece about Aristotle's *De Generatione Animalium*.

As I wrote the article, I considered ways to write links into it so that it would be part of the semantic web that was forming among the entries in the Encyclopedia. While Aristotle's description of reproduction in *De Generatione* was certainly influential in the history of embryology, it was important to trace how it was influential in a meaningful series of connections. I decided to include a short paragraph that described the influence of Aristotle's work on St. Thomas Aquinas, St. Augustine, and St. Albertus Magnus, which in turn linked his work to the Catholic Church. I also described the influence of Aristotle's work on Hieronymus Fabricius and William Harvey, both of whom adopted a revised account of Aristotelian epigenesis. The inclusion of these two paragraphs permitted me to tie Aristotle's work into the history of embryology in a reality-based set of semantic connections.

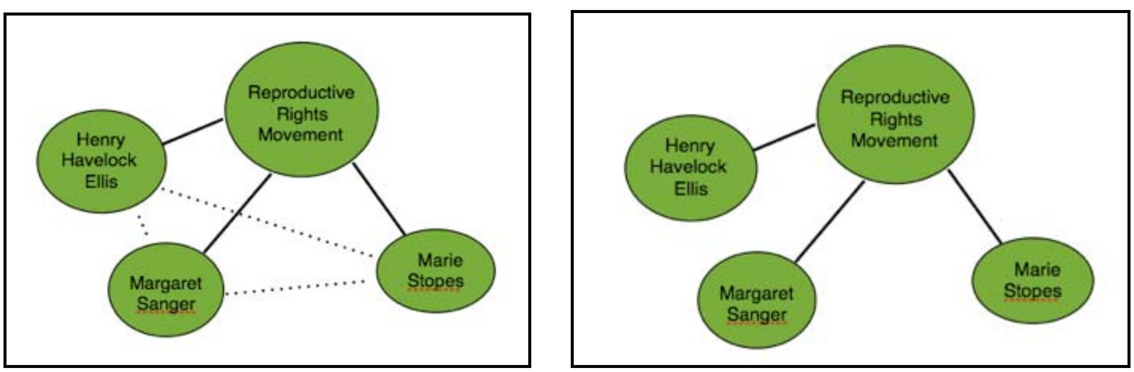
I found myself increasingly interested in the way the "relationship system" worked. With the approval of my PI, I was trained in the process to encode the final edited articles in the XHTML format to upload them to the website (see Appendix C, Chart 1). This process is partially the tedious work of copying and pasting the text of the article, and hand-coding in formatting markup. But doing the markup work also afforded me the opportunity to work directly with the Fedora system. I used the repository search page to locate PIDs of objects that would be linked from the entry in question (see figure "screen grab of search repository"). As shown in Figure 3.1, the PID of the linked object is used as the `href=` in the anchor tag, and the `class=` of the anchor tag asserts the relationship. If an object was not found, I decided what type of object it was and entered it as a placeholder.

The most interesting part of doing the markup work was the actual decision-making process of which class of relationship was most appropriate for any given link. The article authors (usually undergraduate researchers) are instructed to suggest the relationship class they think most appropriate in brackets next to the word that is to be linked. However, the article author's first priority is to create a well-researched and well-written article, and the assignment of semantic relationships requires a bit of mental gymnastics that are difficult to get used to. The relationship suggestions, therefore, are often not quite correct[5]. The markup editor, as the final step in the posting process, has

[5] Though they are sometimes amusingly editorial, as in the case of the suggested relationship between Henry Havelock Ellis and his openly lesbian wife Edith Mary Oldham Lees:

the task of referring to the ontology sheet and making the final call about which relationship actually applies. For me, this involved a lot of thought about the shape of the semantic web that was being created. I wanted to be very careful to assert relationships only if they represented real-world connections, and not merely for the sake of making the link.

Checking the ontology sheet indicated that `hasCoWorker` was the closest fit. The ontology sheet elaborates that `hasCoWorker` means "worked with, collaborated with, were co-workers." But while Sanger, Stopes, and Ellis certainly corresponded with each other and were aware of and supportive of each others' activities, they were not co-workers in the sense that one would use the word to describe a relationship today. What ought the relationship web be, in order to provide meaningful information about real-world connections?



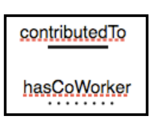


Figure 4.2: The two possible options for the relationship set for Sanger, Ellis, and Stopes. Option 1 provides more information, but does it represent reality accurately?

To make this decision for the Ellis and Stopes articles, I referred to the choices

---

"hasTotallyDysfunctionalMarriageTo". This suggestion was not preserved in markup. In this case, Edith Lees did not seem important to the history of embryology, so no relationship was actually record. This may present a problem, however, for a married couple such as geneticist Charles Davenport and his embryologist wife Gertrude Davenport. Should their relationship be preserved as "hasCoWorker" without recording their marriage? Ontology questions such as this will continue to arise.

made by Dr. Yamashita, who had previously done the XHTML coding for the Margaret Sanger article. Where Sanger "befriended Havelock Ellis," Yamashita had asserted the relationship `hasCoWorker`. Subsequent decisions about which relationship to use in such marginal cases were also made by referring back to similar scenarios in previously published articles, with consideration of how well the relationship described the real-world information.

The work of making relationship decisions eventually lead to an awareness of a gap in the Encyclopedia that ought to be filled. Ellis, Sanger, and Stopes were all listed as supporters of the eugenics movement, with a relationship link asserted to the object "eugenics." An article about early twentieth century embryologist E. G. Conklin also mentioned that researcher's connection to the eugenics movement, and his membership in the Galton Society. Still another article, this one about the Papal Encyclical "Casti Connubii" by Pope Pius XI, linked directly to the object "eugenics" with the relationship `contributedTo`, because that document had condemned eugenic sterilization just as it condemned abortion and contraception.

It seemed that a critical mass of articles concerned in part with the Eugenics Movement of the early twentieth century had accumulated, and to leave that gap unfilled would compromise the utility of the Encyclopedia. The relationship of all of these people, organizations, and writings created clear opportunities to ask questions about what the Eugenics Movement was, who was involved in it, and what influenced it. Most importantly of all, it indicated that the Eugenics Movement was inextricably tied into the history of embryology. Agents from three different social contexts, each important to the history of embryology, were linked to this single object. This indicates that a proper discussion of the Change Agents in the history of embryology cannot take place without including articles about the Eugenics Movement.

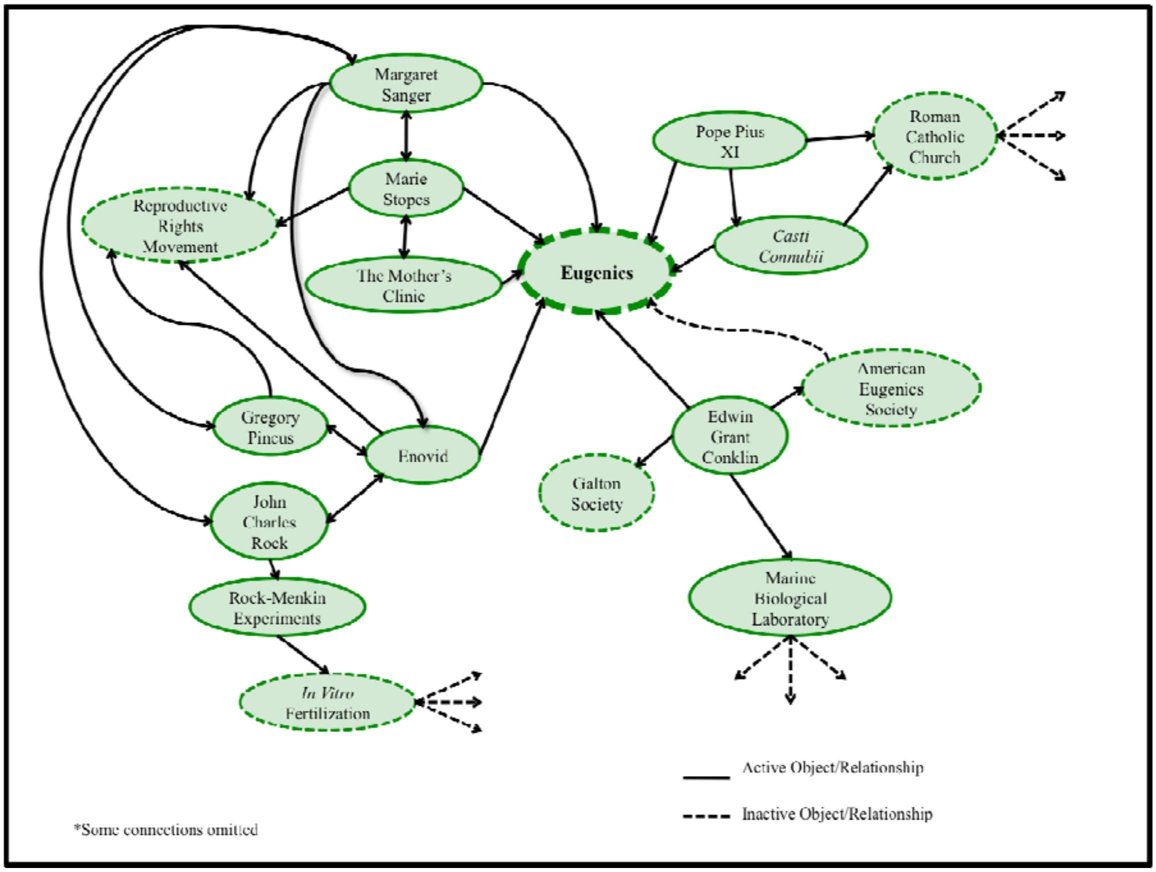


Figure 4.3: A semantic web centered on the object "Eugenics."

This result can immediately feed back into the Embryo Project content development process. The existence of this cluster raises several questions: Why are eugenics and reproductive rights so embedded in each other's histories, and in the history of embryology? How are these things related, and what aspects of the scientific context drove the eugenics and reproductive rights movements? What aspects of those social contexts drove genetics and developmental biology? This could provide fertile ground for a new undergraduate cluster of articles, or (if the questions are addressed with greater sophistication) a graduate level project. In this way, the semantic web technology fosters a collaborative, multidisciplinary approach to the history of science.

It is important to note that the diagram in Figure 4.3 was produced manually in PowerPoint, and was derived from the information encoded in the XHTML files. The diagram of this web does not display the richness of the semantic relationships, but only displays linkage. The semantic relationships that the arrows represent encode such information as "Marie Stopes `contributedTo` Eugenics", "Edwin Grant Conklin `isEmployee` of the Marine Biological Laboratory," "John Charles Rock `isInventor` of Enovid," just to describe a few examples. These relationships are actually indexed and stored in the Fedora repository as an RDF datastream. From there, the relationships can be queried, used to refine searches, used to guide a user through a predetermined path in the Encyclopedia, and used to create dynamic diagrams that a user can interact with in multiple ways (Lagoze 2006). Some development has already been completed along these lines, but more is needed.

5. Opportunities for the Future

There is a lot of hype about "Web 2.0" these days. Everyone seems to be talking about how the Internet will completely change the way people approach information, education, and communication.[6] It would be worthwhile to examine how much this is really the case, but for now I am simply adding my voice to the consensus. In The Embryo Project, using the semantic web completely changes—for the better—the way in which researchers can create, display, and process their results. The admittedly modest

[6] A sample of blog headlines: "Web 2.0 is Changing the Way We Work" (http://www.theappgap.com, 2008), "Web 2.0 Changes the Way We Connect" (http://www.fiercecio.com 2007), and "Why Web 2.0 Changes The Way IT Works" (http://in-cider.spaces.live.com 2008).

case study of a cluster of articles related to eugenics is an example of this, in which the use of semantic web technology has revealed gaps in the work that has been done in the Embryo Project. But there are other ways in which one can envision this sort of technology really adding something new to the research and presentation of history of science.

The Embryo Project could function as an entirely new sort of reference work—a hybrid of the engaging, narrative format of popular history books and the highly informative and easy-to-use encyclopedia format. Clara Pinto-Correia’s book *The Ovary of Eve* is an illustrative example of the first format. One of the strengths of Pinto-Correia's book, besides her engaging writing and strong research, is that she presents the history of the Preformationism theory in terms of what seems to have influenced the scientists involved, and describes how Preformationism may have influenced the culture around it. This makes the history more interesting to read, and helps explain to a modern reader why such brilliant scientists could have been so wrong about their theories. This complex story about how the scholars involved got to their conclusions is useful to present when dealing with defunct theories such as Preformationism, or the Phlogiston Theory of Combustion, or Geocentricism. Modern students of history, especially the younger ones, may not be fully aware of their own privileged frame of reference. They need to be guided through the past scholar's process (Griesemer 2007).

But while Pinto-Correia does an excellent piece of work in describing the influences and processes of the scientists she writes about, the book format is still restrained to a linear telling. She is made to circle back narratively, and explain a scientists' past after she introduces him to the reader, or take a side-trail to explain a coworker's influence on a certain piece of work. If one is reading her work as a set piece, this is engaging and enjoyable; good historians are usually good storytellers. But as a work of reference, Pinto-Correia's history is unusable. It is not possible to open to her discussion of one scientist or one experiment and get the full story at once. Nor is that the purpose of her piece. For the full story of one particular actor, one uses a reference work.

A reference work like the *Dictionary of Scientific Biographies* fulfills its function quite well, of course, but it ca not provide the elegance and explanatory power of a good narrative history. The beauty of the Embryo Project Encyclopedia, as it might be, is that it can do the job of both. By being composed of discrete entries and media objects, it is searchable like a reference work and can give a full overview of any item in the history of embryology. But because semantic relationships are recorded as the entries are published, the database is full of the information that a narrative history provides: influences, coworker relationships, correspondence, organization memberships, mentorship, and a slew of others. The main drawback has been that it simply takes time to create that web of relationships, because each entry has to be written and linked up to others. The Embryo Project has been in progress for two years now, and only now that there is a critical mass of entries can we begin to look at the connections.

While the work of content creation must certainly continue, a second phase of the project appears to be taking shape. That work will involve creating the end-user experience, and making several different engaging and intuitive ways in which to explore the content at multiple levels. Accurate and reliable searching for one particular piece of information is only one of these ways. Online exhibits and dynamic diagrams of the semantic webs are other possible ways. However, the creation of these web applications is well outside the focus of the Embryo Project's NSF grant, and even with the support of the ASU Library, would probably not be possible in a reasonable time frame. If the Embryo Project had to create these things on its own, the project might have had serious obstacles to achieving real social relevance beyond the academic sphere.

Luckily, the choice of Fedora as the software tool means that the Embryo Project is a member of a community of similar project groups who are working together to create the kinds of end-user tools needed. The Fedora software is produced by a non-profit organization called Fedora Commons, which partners closely with several digital repository projects, including the National Science Digital Library[7] (NSDL) and the Max Planck Society, and the Public Library of Science[8] (Payette 2007). The NSDL provides copious online resources to science educators at all levels, from elementary school to university. The Max Planck Society's *eSciDoc*[9] is a free repository and collaboratory for science data—in other words, a place for researchers to upload not just their publishable results, but content relating to "all intermediate steps as well" (eSciDoc 2008). The Public Library of Science serves as a free and public scientific journal, a resource considered by many to be of immense social importance (Lessig 2006). These and a few other projects represent the main partners who lead the development of Fedora source code and Fedora implementation, but there are 156 known projects in all that use Fedora and participate in development (FedoraCommons 2009).

The Embryo Project can also look to other projects that use Fedora to get ideas about how we might format the user experience web applications. The Encyclopedia of

[7] http://nsdl.org/
[8] http://www.plos.org/
[9] http://www/escidoc.org/

Chicago[10] is a repository for archived information similar to the Embryo Project, and in addition to having a simple searchable encyclopedia of entries pertinent to Chicago, it hosts many special features such as "Digital Interpretive Essays" that guide users through several encyclopedia entries at a time. The Biodiversity Heritage Library[11] presents a "tag cloud" on its home page, which provides an instant view of what kinds of articles it contains, and which kind are the most numerous. The Encyclopedia of Life[12], an ambitious project to catalogue "virtually all information about life present on Earth," provides a brief video guide to using its website (though its interface is remarkably intuitive). Through their connections with the MBLWHOI Library, Max Planck Laboratories, and the Fedora Commons organization itself, the Embryo Project and ASU Libraries can adopt and adapt many of these presentation practices without having to reinvent them all over again.

A more formal aspect of the Embryo Project's collaboration with other Fedora-based projects involves an upcoming project to share data across domains. This collaboration will result (it is hoped) in each group's database having semantic web information pertaining to the other groups' databases. The development of a shared ontology—that is, an agreed upon common set of relationship classes—will enable all of these projects to access and index each other's information. Automatic annotation software will be able to use the shared ontology to perform the tedious work of assigning these relationships, increasing the speed at which this information is produced (Royer, et al. 2007). Joint research from the Embryo Project and the Marine Biological Laboratory

[10] http://www.encyclopedia.chicagohistory.org/
[11] http://www.biodiversitylibrary.org/
[12] http://eol.org/

has already produced "proof of concept" work of the annotation and indexing process. Using several different ontologies, each suitable for a different domain (and thus a different type of project), the team scraped 1000+ Medline abstracts and mined them with an annotation application to detect various information (Yamashita, personal communication) . Though still at its infant stage, this ability to mine text will not only find objects of interest, but the semantic relationships of these objects to other objects. The semantic web represented by the eugenics cluster is only a small sketch of the possibilities presented by such an enormous collection of semantic relationships.

Having made Lego-like models of their own particular areas of interest, these separate institutions can now join their models together at the corners to make a Lego super-model. The World Wide Web Consortium (W3C) puts it very elegantly on their page introducing semantic relationships:

> "The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources….It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing." (Herman, et al. 2009)

The Embryo Project, on its own, has shown that it is capable of creating a semantic web that is useful to the historian of science. We can see different stories about the progression of history when we use semantic web technology, and we are able to tell different stories, and tell the stories we have told differently. The fact that the semantic relationship technology can combine the Embryo Project's knowledge with other databases in ways that can only hazily be predicted means that the researchers in the Embryo Project—undergraduates, graduate students, and experienced researchers—are participating in an interdisciplinary collaboration that is perhaps unprecedented. Even if the Embryo Project were an island of academia, entire of itself, the use of semantic relationship technology to view and organize semantic webs within its content would be of great value in the identification of Agents of Change.

Bibliography

Allemang, Dean, and Jim Hendler. *Semantic Web for the Working Ontologist.* Burlington, MA: Morgan Kaufmann Publishers, 2008.

Baker, Christopher J. O., and Hei-Hoi Cheung, . *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences.* New York, NY: Springer Science+Business Media, 2007, 2007.

Brind'Amour, Katherine. "Casti Connubii." *Embryo Project Encyclopedia.* 2009. http://embryo.asu.edu/encyclopedia/literature/?getObject=embryo:124885 (accessed February 21, 2009).

Buettner, Kimberly A. "John Charles Rock." *Embryo Project Encyclopedia.* 2007. http://embryo.asu.edu/encyclopedia/people/?getObject=embryo:124783 (accessed February 21, 2009).

—. "Margaret Higgins Sanger." *Embryo Project Encyclopedia.* 2007. http://embryo.asu.edu/encyclopedia/?getObject=embryo:124756 (accessed February 21, 2009).

Codd, Edgar F. *The Relational Model for Database Management Version 2.* Boston, MA: Addison-Wesley Longman Publishing Co., Inc., 1990.

Creative Commons Organization. "Creative Commons Legal Code." *Creative Commons.* http://creativecommons.org/licenses/by-nc-nd/3.0/us/legalcode.

Dupont, M. Ellen M. "Henry Havelock Ellis." *Embryo Project Encyclopedia.* Edited by Jane Maienschein. 2008. http://embryo.asu.edu/encyclopedia/?getObject=embryo:124866 (accessed February 21, 2009).

—. "Marie Charlotte Carmichael Stopes." *Embryo Project Encyclopedia.* Edited by Jane Maienschein. 2008. http://embryo.asu.edu/encyclopedia/?getObject=embryo:125263 (accessed February 21, 2009).

eSciDoc. *eSciDoc: Overview .* eSciDoc. August 5, 2008. http://www.escidoc.org/ (accessed March 19, 2009).

Griesemer, James. "Tracking Organic Processes: Representations and Research Styles in Classical Embryology and Genetics." In *From Embryology to Evo-Devo: A History of Developmental Evolution*, edited by Manfred Laubichler and Jane Maienschein, 375-433. Cambridge: The MIT Press, 2007.

Gruber, Tom. "A Translation Approach to Portable Ontology Specifications." *Journal of Knowledge Acquisition* (Academic Press) 5, no. 2 (1993): 199-220.

Gruber, Tom. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing." *International Journal of Human-Computer Studies* (Academic Press), no. 43 (1993): 907-928.

—. *What is an Ontology.* http://www-ksl.stanford.edu/kst/what-is-an-ontology.html (accessed March 6, 2009).

Harnad, Steven. "When "Knowledge Engineers" say "Ontology" They Mean The Opposite: "Epistemonomy"." *Skywritings.* December 20, 2008.

http://users.ecs.soton.ac.uk/harnad/skywritings/index.php?/archives/82-guid.html (accessed March 6, 2009).

Herman, Ivan, Sandro Hawke, Eric Prud'hommeaux, and Dave Raggett. "W3C Semantic Web Activity." *World Wide Web Consortium.* February 26, 2009. http://www.w3.org/2001/sw/ (accessed March 8, 2009).

Hofweber, Thomas. "Logic and Ontology." *Plato: The Stanford Encyclopedia of Philosophy.* October 13, 2004. http://plato.stanford.edu/entries/logic-ontology/ (accessed March 6, 2009).

Lagoze, Carl, Sandy Payette, Edwin Shin, and Chris Wilper. "Fedora: An Architecture for Complex Objects and Their Relationships." *International Journal on Digital Libraries* (Springer) 6, no. 2 (April 2006): 124-138.

Laubichler, Manfred, and Jane Maienschein. "Project Updates: Understanding Agents of Scientific Change: The Case of Embryo Research." *2007 HSD Grantees Conference Proceedings.* http://www.nsf.gov/sbe/hsd/2007.pdf. 162.

Lessig, Lawrence. *Free Culture.* 2006.

Lubyte, Lina. "Reusing Relational Sources for Semantic Information Access." *PIKM '07.* Lisbon, 2007. 1-8.

Maienschein, Jane. "Why Study History for Science?" *Biology and Philosophy* (Kluwer Academic Publishers), no. 15 (2000): 339-348.

Maienschein, Jane, and Manfred Laubichler. "The Embryo Project Proposal to the NSF Human Dimensions." *Grant Proposal.* 2006.

Paul, Diane B. *Controlling Human Heredity: 1865 to the Present.* Edited by Margaret C. Jacob, Spencer Weart and Hal Cooke. Amherst, New York: Humanity Books, 1998.

Payette, Sandy. "Fedora Commons Proposal to the Gordon and Betty Moore Foundation." *Fedora Commons Organization.* 2007. http://www.fedora.info/pdfs/Fedora_Commons_Proposal_to_the_Moore_Foundation.pdf.

Pinto-Correia, Clara. *The Ovary of Eve: Egg and Sperm and Preformation.* 1st Edition. Chicago: University of Chicago Press, 1998.

Royer, Loic, Benedikt Linse, Thomas Wachter, Tim Furch, Francois Bry, and Michael Schroeder. "Querying Semantic Web Contents: a Case Study." In *Semantic Web: Revolutionizing Knowledge Discvery in the Life Sciences*, edited by Christoper J. O. Baker and Hei-Hoi Cheung, 31-52. New York, NY: Springer Science+Business Media, 2007.

Ruffenach, Stephen C. "Rock-Menkin Experiments." *Embryo Project Encyclopedia.* 2009. http://embryo.asu.edu/encyclopedia/experiments/?getObject=embryo:125740 (accessed February 21, 2009).

Schuermann, D. Brian. "Edwin Grant Conklin." *Embryo Project Encyclopedia.* 2009. http://embryo.asu.edu/encyclopedia/people/?getObject=embryo:123965 (accessed February 21, 2009).

Storey, Veda C. "Understanding Semantic Relationships." Edited by Fred J. Maryanski. *VLDB Journal* (Springer-Verlag) 2, no. 4 (October 1993): 455-488.

The Plant Ontology Consortium Project. *Plant Ontology (PO).* January 21, 2009. http://www.plantontology.org/ (accessed March 6, 2009).

Appendix A: EP Ontology Sheet

Object 1	Object 2	relationship expression	relationship definitions
PEOPLE	People	hasTeacher	student of, taught by, educated by, advised by
		hasStudent	had students, taught, educated, supervised
		hasCoWorker	worked with, collaborated with, were co-workers
		hasCoAuthor	co-authored with, collaborator, co-edited
	Organization	isStudent	studied at, educated at, visited, was a student at
		isDirector	was head of, chaired, directed, president of
		obtainedMD	obtained an MD from, got MD from, received MD from
		obtainedPhD	obtained a PhD from, got PhD from, received PhD from
		obtainedDegree	obtained degree from, got degree from, graduated from
		isEmployee	worked at, hired at, employed by
		isResearcher	did research at, visited, conducted work at; the difference between this and isEmployee is that the person was not officially hired or employed by the organization
		fundedBy	obtained funding from, got grants from, was awarded fellowship
	Place		
	Experiment	conducted	conducted, performed, designed, developed, carried out
	Technology	isInventor	invented, designed, developed, created, originated
		used	used, experimented with, manipulated
	Organism	isInventor	invented, designed, bred, manipulated
		used	used, experimented with, manipulated
	Law	hasRelevance	is relevant to,
	Religion	hasAffiliation	was affiliated with, funded by, joined
	Ethics	hasRelevance	is relevant to,
	Scientific Context	isEmbeddedIn	
		contributedTo	

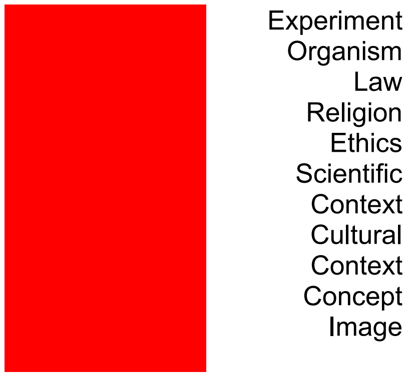
	Cultural Context	isBackgroundTo isEmbeddedIn contributedTo	
	Concept	isBackgroundTo contributedTo	contributed to, defined, articulated, explained, developed, modified
	Image	isCreator	created, painted, drew, photographed, diagrammed, developed, recorded, filmed, sculpted
		used	used, included, stole, adapted, published, copied, commented on, referred to
		depictedIn	is depicted in, is shown in, illustrates, represents
	Literature	isAuthor	wrote, co-wrote, penned
	Award	hasAward	won, was awarded, was presented with, was honored with
PLACE (geographic entity, location)	People		
	Place		
	Organization		
	Experiment	isLocation	site of, conducted at
	Technology		
	Organism		
	Law		
	Religion		
	Ethics		
	Scientific Context		
	Cultural Context		
	Concept		
	Image	depictedIn	
	Literature		
	Award		
ORGANIZATION (institution, association, etc.)	People	hasGraduate	awarded a degree to, graduated, educated
		funded	funded, gave grant to, awarded fellowship to

		fundedBy	funded by
		hasEmployee	employed, hired
		hasDirector	had as a director, president, leader, etc.
		hasMember	had as a member
		hasEditor	had person that served in editorial role
		hasInstructor	employed as instructor or teacher
	Place	hasLocation	
	Organization	funded	
		fundedBy	
		isRelatedTo	
		hasMember	
		hasAffiliation	is affiliated with
	Experiment	conducted	conducted, performed, designed, developed, carried out
		isPublisher	published
	Technology	isInventor	invented, designed, developed, created, originated
	Organism	isInventor	housed, bred
		used	invented, created, designed
	Law		
	Religion	hasAffiliation	affiliated with
	Ethics		
	Scientific Context		
	Cultural Context		
	Concept		
	Image	isPublisher	
	Literature	isPublisher	
	Award	hasAward	was awarded, won, was given
CONCEPT	People	usedBy	used by, rejected, modified, propagated, invented, contributed to
	Place		
	Organization	usedBy	used by, discussed in, developed by, adapted by, rejected by
	Experiment	isBackgroundTo	background to, contributed to
	Technology	isBackgroundTo	background to, contributed to
	Organism		
	Law	isBackgroundTo	background to, contributed to
	Religion	hasRelevance	relevant to, background to, contributed to
	Ethics	hasRelevance	relevant to, background to
	Scientific Context	isEmbeddedIn	
		contributedTo	
		isBackgroundTo	
	Cultural Context	isEmbeddedIn	
		contributedTo	
		isBackgroundTo	

	Concept	contributedTo	
	Image	depictedIn	depicted, illustrated, demonstrated, captured
	Literature	discussedIn	
	Award		
IMAGE	People	hasCreator	created by, painted by, drawn by, photographed by, diagrammed by, developed by, recorded by, commissioned by
		depicts	
	Place	depicts	depicts, shows, illustrates,
	Organization	locatedIn	depicted in, published in, housed in
		hasCreator	
		depicts	
EXPERIMENT	People	conductedBy	conducted by, did, tried, performed, developed, designed, carried out
	Organization	conductedBy	conducted by, performed, done, developed, designed, carried out
		fundedBy	funded by, endowed by, supported by
	Place	conductedAt	
	Technology	used	used, utilized
	Organism	used	used, killed, ablated, manipulated, bred, raised, cloned
	Law	hasRelevance	is relevant to, is background to
	Scientific Context	isEmbeddedIn	
		contributedTo	
		isBackgroundTo	
	Cultural Context	isEmbeddedIn	
		contributedTo	
		isBackgroundTo	
	Concept	contributedTo	contributed to, led to
	Image	depictedIn	is depicted in, shown in, drawn in, illustrated in, photographed in
	Award	contributedTo	
TECHNOLOG Y	People	hasInventor	invented by, designed by, created by, developed by, originated by, crafted by
	Organization	hasInventor	invented by, designed by, created by, developed by, originated by, crafted by
		hasFunding	funded by, financed by
	Place		
	Experiment	contributedTo	contributed to
	Technology	contributedTo	contributed to
		resultedFrom	resulted from; followed;
		relatedTo	related to; similar to;

	Organism	used	used, killed, manipulated
	Law	hasRelevance	is relevant to
	Religion		
	Ethics	hasRelevance	is relevant to
	Scientific Context	isEmbeddedIn	
		contributedTo	
		isBackgroundTo	
	Cultural Context	isEmbeddedIn	
		contributedTo	
		isBackgroundTo	
	Concept	contributedTo	contributed to
	Image	created	created, originated, developed, invented, produced
		used	used, manipulated, altered, copied, magnified, clarified
		depictedIn	is depicted in, is shown in, is photographed in, is drawn in, is illustrated in
	Award	contributedTo	
LITERATURE	People	hasAuthor	
	Place		
	Organization		
	Experiment		
	Organism		
	Law		
	Religion		
	Ethics		
	Scientific Context		
	Cultural Context		
	Concept	hasTopic	
	Image		
ORGANISM	People		
	Place	locatedIn	
	Organization	usedBy	
	Experiment	usedBy	
	Organism		
	Law		
	Religion		
	Ethics		
	Scientific Context		
	Cultural Context		
	Concept		
	Image	depictedIn	is depicted in, shown in, drawn in, illustrated in, photographed in
	Award		
LAW	People	hasAuthor	
		hasLitigant	comprises those people who are not only directly

			mentioned in the case title (e.g. ÒRoeÓ or ÒWade,Ó but also those who are major players in the case
	Place		
	Organization	hasLitigant	
	Experiment		
	Organism		
	Law	precedes	the case/statute article of interest is a precedent for this later case/statute/law; the case of interest is cited in this later case
		reliesOn	the case/statute article of interest relies on this previous case/statute/law; the case mentions this earlier case/law
	Religion		
	Ethics		
	Scientific Context		
	Cultural Context		
	Concept	defines	the ruling of the case in effect defines this concept
		isConcernedWith	the case relies on these concepts for much of its argumentation, but the case itself does not define these concepts; the topical matter of the case
		mentions	the case mentions these concepts and has loose relevance to these concepts. weaker relationship than ÒdefinesÓ or ÒisConcernedWithÓ
	Image		
RELIGION	People		
	Place		
	Organization		
	Experiment		
	Organism		
	Law		
	Religion		
	Ethics		
	Scientific Context		
	Cultural Context		
	Concept		
	Image		
ETHICS	People		
	Place		
	Organization		


Experiment
Organism
Law
Religion
Ethics
Scientific
Context
Cultural
Context
Concept
Image

Appendix B: Process Flow Charts

Chart 1: XML Coding Process Map. Notice the decision in the middle of the flow, "Does appropriate relationship class exist for this link?" This is currently a subjective decision made with guidelines, but future development of data mining processes can automate the relationship class decision.

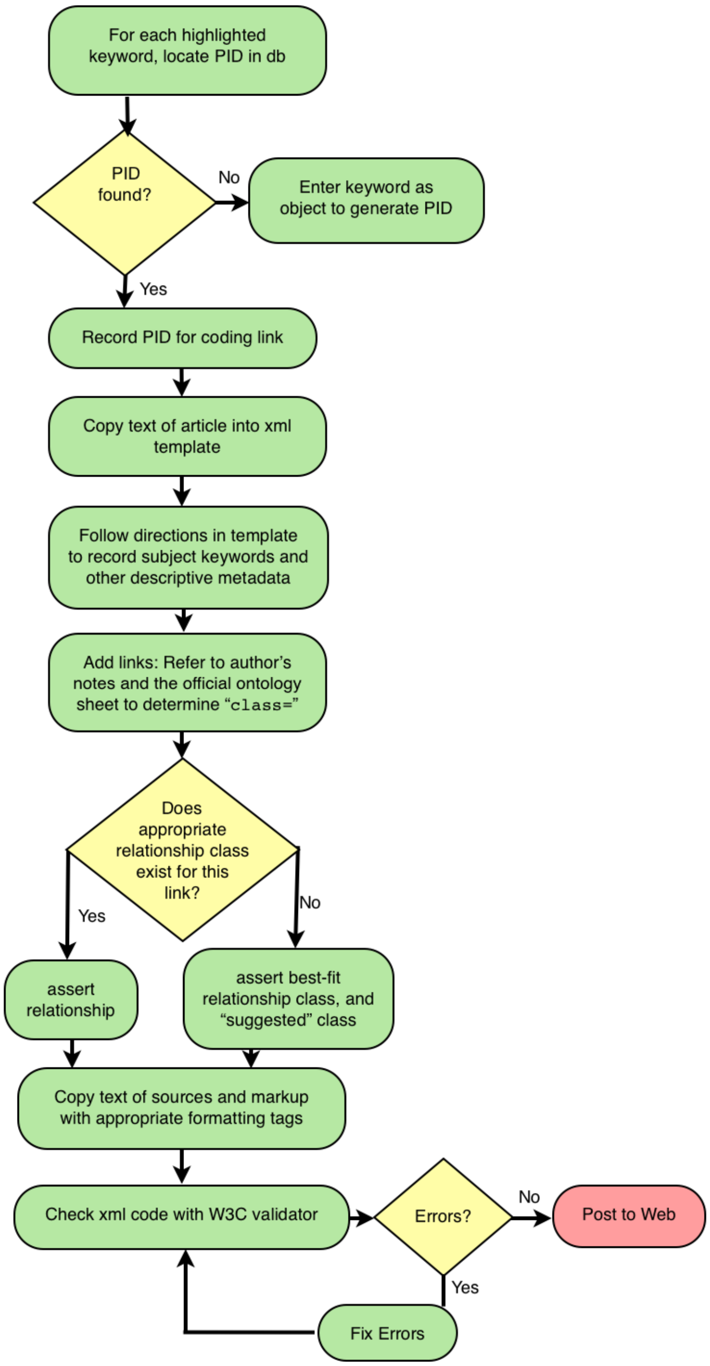


Chart 2: Undergraduate Entry and Relationship Process. This describes the steps that undergraduate researchers take when drafting an entry, with emphasis on the review and feedback steps.