# Implementing a Born-Digital Architecture and Design Archive at the Art Institute of Chicago

Carissa Kowalski Dougherty
Art Institute of Chicago
111 S. Michigan Avenue
Chicago, IL 60603
ckowalski@artic.edu

**Abstract**

This presentation will provide an overview of the issues confronted by a small team of archivists, architects, and programmers in developing a pilot system for collecting, archiving, and exhibiting born-digital design data in the Department of Architecture and Design at the Art Institute of Chicago. The pilot is the culminating activity in an 24-month implementation program; its final products include a functioning software ingest system (Digital Archive for Architecture, or DAArch), a workflow procedure for the curatorial department, and a small collection of accessioned, cataloged, and preserved digital architecture and design objects. As the Department's coordinator for the digital archiving project, I will be presenting the findings of our pilot from the perspective of someone working within large visual arts institution, negotiating among several interest groups, and addressing a wide range of problems, from copyright issues to metadata schema, from format standards to system architecture. Using the workflow diagram developed for the pilot project as a starting point, I will trace a typical born-digital collection object as it is ingested and processed using the DAArch system. This narrative will focus on the challenges we faced in developing something conceptually and technologically complex as well as completely unprecedented.

**Implementing a Born-Digital Architecture and Design Archive at the Art Institute of Chicago**

## Background and Overview

A curatorial department of architecture within a fine arts museum faces many unique challenges. Architecture as a discipline negotiates the fine line between fine and applied arts. A skilled architectural rendering, for example, may be suitable for displaying as a work of art alongside a Monet—but, of course, many architectural outputs are practical in nature, intended as instructions for builders or contractors. This is not to say that plans and sections do not have a beauty unto themselves; it only suggests that the type of curatorial discretion that happens in, say, a department of European painting is not quite the same when applied to architectural representations, where pieces are chosen not only for their artistic merit, but for their value as records of the design process. Architecture is also interesting because what we collect is—for the most part—only a representation of the final product, which is the building itself. While certainly some museums have "collected" works of architecture, such as period rooms or, in the example of MoMA's PS1 program, commissioned installations, it is virtually impossible to collect actual buildings. And so we collect the drawings, models, and other forms of representation instead.

But rapid changes in media used in architectural representations present an even greater challenge. The late twentieth century saw a rapid increase in the use of digital tools in architecture firms. Where computer-aided drafting (CAD) software was once simply a more efficient way to product construction documents, there are now hundreds of software programs that allow architects and designers to create three-dimensional models, lifelike digital renderings, and even animations of their projects. These "born-digital" products of the design process are essential records of contemporary architectural practice.

But very few born-digital documents ever make it to hard copy, nor were they intended to be printed. In fact, for those projects that never get built—such as competition entries or conceptual work—there may be little "paper trail" at all. Right now, the architecture collection at the Art Institute includes only paper drawings and physical models. What if a contemporary practitioner wanted to donate their latest born-digital project? Does it really make sense to print out digital renderings, to "freeze" animations and mount them on the wall? Perhaps not, but the obstacles to collecting digital work are numerous, as most of you here at this conference know. The sheer number of software programs used even within a single architecture firm is staggering. Both software and hardware quickly become obsolete; even within a firm, it is possible for designers to lose access to their own work, as they allow their software licenses to expire, turn over staff, or simply move on to a different medium. Without a systematic way of preserving these digital design documents, we stand to lose a great deal of history. And, without a way to preserve design intent in a medium approximating the original, we also stand to lose a great deal of information about the digital design process itself.

In order to address these issues, the Department of Architecture and Design at the Art Institute of Chicago, in collaboration with Kristine Fallon Associates, undertook a study to find out what it would take to collect, archive, and exhibit born-digital design data. The study included several steps of research and validation, including: (1) a survey of over a hundred design firms worldwide, inquiring about their use of different types of software; (2) in-depth case studies to investigate how digital tools are used in nine firms' typical design processes; (3) an examination of best practice standards for archiving digital materials in museums, libraries, and other cultural heritage institutions; and (4) brainstorming sessions with experts from several fields, including imaging,

library science, and software technology. The specific findings of this study are beyond the scope of this presentation; however, many of the recommendations from the study form the basis of the pilot project I will be discussing today. For those of you who are interested, the final report is available online in pdf format at www.artic.edu/aic/collections/dept_architecture/ddd.html or you may come up after the presentation to pick up one of the CDs.

In 2004, we received funding to begin a real-world implementation of the study. The goals of the 24-month pilot project were to develop a working software cataloging/ingest system, establish a workflow procedure for collecting digital materials, and produce a small archive of fully cataloged and preserved digital materials. Of course, an actual implementation always brings out new issues that were not anticipated in the more conceptual, research-based study. As the digital archive coordinator for this project, I will be talking about some of the issues we confronted in this pilot project, and presenting some of the outcomes of our work. As a starting point, I will use the workflow diagram adapted for the study, based on the Open Archival Information System (OAIS) reference model (Figure 1). Following the steps in this diagram, I trace the process of archiving digital materials as envisioned in this pilot project.

**Step 1: Preparing Data for Submission into the Archive**

Preparing data for submission into the digital archive involves the curator visiting the architect's office and working with them to determine what files they would like to collect. This step is the most exciting, but may also be the most frustrating; architecture firms vary so widely in their file naming, organizational, and workflow procedures that it is extremely difficult to impose any kind of consistent standards.

In the original study, we recommended a hierarchical, highly codified system of organization called the National CAD Standard. For large, corporate firms, the National CAD Standard or a similar naming convention is feasible. They have the organizational resources to accomplish high levels of organization, and may have internal libraries, archives, or marketing departments that maintain efficient and logical systems for naming and organizing project files. The reality is, however, that smaller firms have neither the staff nor the time to maintain consistent standards or even keep track of the vast number of files generated for an architectural project. At a minimum, we have been asking firms to organize their data in a way that is consistent and transparent. One logical structure is to place files in directories named by project phase and drawing type, as we see here (Figure 2).

One of the most important aspects of the archiving workflow is our use of a two-tiered approach to collecting born-digital design data (Figure 3). The first tier is the "native data"—the original files, in the formats in which they were created. As mentioned previously, files for any given project may exist in dozens of different formats, from common software like Adobe Photoshop to more specialized software such as CATIA. While we don't necessarily expect these files to be accessible in a matter of years (or even months), preserving at least the basic bitstream allows for the possibility of researchers coming in and recreating or emulating the original platform at a later date. Parallel projects in the IT world, including the creation of format registries, make digital archaeology a realistic and essential discipline for the twentieth century. Keeping these data is relatively inexpensive, in terms of servers and storage media, and it is much less costly than losing the data altogether.

The second tier is the "output data"—the files selected from the larger pool of native data that are identified by the architect or curator as being significant milestones in that project's design

process. Based on the variety of tools we found in the profession, we needed to find formats that would remain relatively stable over time. One way to ensure this is by using non-proprietary, un-encrypted formats that are widely accepted as standards and easily writable from a variety of design software. We have found that tagged image file format (or tiff) and portable document format (or pdf) best meet these needs. Both of these formats allow for color profiles to be stored as a part of the file, and can be saved without lossy compression.

Rendered images and digital photographs are saved as tiffs, at a high resolution. Line drawings, such as floor plans or other construction documents, are saved as pdfs. The latest versions of Adobe Acrobat and some CAD software allows for layers, line weights, and even views to be captured and preserved in the pdf format. Three-dimensional models may also be captured as pdfs, although this mechanism is fairly new and has not been widely tested with the entire range of modeling software. At the very least, it is important to be able to interact with the design object. The development of Building Information Model (BIM) software is also pushing archivists to find a solution to three-dimensional design output that will allow future viewers access to the data. Ideally, we would collect files from both of these tiers; this approach allows curators to place higher priority on those files he or she deems most relevant and worthy of being accessioned as "art" objects.

Another set of issues that arise when dealing with donors of digital files are those involving copyright and the idea of a digital "original." When plans and renderings are generated as digital files, they may be copied at will, with no reference to an "original." Additional changes may be made to these files by people other than the so-called "author"—and this collaboration is in fact often encouraged in contemporary practices. It has been difficult to educate the museum's administrative and legal departments about these issues, especially given the premium that is placed on the value of an original piece of art in a museum setting. The solution that we've come up with is a document called a "Non-Exclusive License for Copyright." In signing this document (which replaces our old Deed of Gift), the copyright owner grants our institution a license for the life of the copyright to reproduce the digital work for exhibitions, publication, and educational purposes. The wording of this document is quite interesting and necessarily ambiguous, allowing us to reproduce the images in "any media now known or not yet invented." The license was seen as better than an outright transfer of copyright for two reasons: One, we don't then have the responsibility of maintaining the copyright ourselves, and two, the creator would be more likely to donate his or her work if he or she is still able to control the copyright for their own purposes.

**Steps 2–4: Processing, Cataloging, and Storing Data**

Once the data has been ingested into the digital archive, metadata is assigned and an Archival Information Package (AIP) is created. This is accomplished through the Digital Archive for Architecture (DAArch) software developed by an in-house programmer, together with consultants from Kristine Fallon Associates. In the beginning of the implementation, the project team determined several requirements that the software must meet in order to succeed as a mechanism for ingesting digital materials. First, the system needed to be easy for a curator or archivist to use; it had to have an intuitive user interface and an efficient input technique. Second was the requirement that the system could be installed at other institutions. This was a stipulation made by our primary donor, who insisted the end product in some way be made available to the archival community and architectural profession, so that it might evolve and remain relevant as technologies change. Third, that the system could stand alone or, alternatively, operate in conjunction with existing collection management systems. Lastly, the system needed to be maintainable over the long term.

Without going into too much technical detail, it is possible to describe the overall structure of the DAArch software with a simple diagram (Figure 4). The system logic is what communicates between the back-end repository, where the data are stored, and the front-end user interface, or client. During the original study, DSpace easily distinguished itself as the most viable option for storing digital design data. This open-source repository developed by MIT and Hewlett Packard to archive academic data is nonproprietary and has become popular among a large user community that share their innovations freely. And, as we will see later, DSpace provides several options for digital preservation. We are also trying to make the program logic as transportable as possible, allowing it to bridge and communicate with existing databases or collection management systems that may already be in place at any given institution. In the case of the Art Institute, we work with an in-house collection management system called CITI, which will eventually take the place of the front-end user interface but still operate with a DSpace repository. In this way, users can gain additional functionality—namely, the ability to "attach" digital files to their metadata—yet retain the more familiar interface.

DAArch is a browser-based software, completely platform-independent—that is, it can be used on a Mac or PC. While I will not go into much detail on any one of the screens, for lack of time and in the interest of keeping everyone awake, I will try to highlight some of the features that have improved upon our current collection management system. In doing so, I hope to explain how some of the design decisions were made and how these decisions reflect the particular needs of the curatorial department of architecture. The examples you will see here are actual catalog records from the archive given to us by Garofalo Architects. This particular project, the Manilow Residence, is significant in that it used software to generate the undulating form of the roof shell and to communicate directly with CNC milling machines that produced the structure of the house.

DAArch, following the recommendation of the original study, is based on the Categories for the Description of Works of Art (CDWA) metadata standards developed by the Getty Institute. CDWA allows for hierarchical data—very important when dealing with part/whole relationships as we do when we're talking about sets of drawings describing a building—and it was developed as a cataloging schema for works of art, so it more closely matched our needs. But we also needed to include architecture-specific fields and to "invent" metadata fields related to digital media. We erred on the side of including too many fields; we knew that anticipating all future uses of the software was impossible, but also recognized that capturing as much data as possible would be especially important for future viewing, research, and exhibition purposes.

The first thing you see is the login screen asking for your username and password (Figure 5). Different read/write permissions are granted to different users based on their role in the archival process. This allows the administrators to control who may add and delete records, for example, and who may only search and view the data. After validating this information, you are taken to a home screen where you have a few different options for entering metadata (Figure 6).

One way to quickly add core metadata is by using the Simple Add screen (Figure 7). The Simple Add screen contains the minimum required fields for an object record. Here, the cataloger would enter title, creator data, drawing type, indexing terms, dates, dimensions, media, reference numbers, etc. This simplified data entry form is best used with groups of native files, or in cases where the cataloger has to quickly enter basic information and intends to flesh them out later.

To enter more comprehensive data, you would navigate to the Normal Add screens (Figures 8–15). The Normal Add tabs are roughly based on the CDWA schema and provide data entry fields for several categories of information. A thumbnail image of the attached file appears at the

top of the screen, along with label text and a navigation bar. A number of features appear on each page that help the user negotiate the data entry fields. To the right of each line are plus and minus symbols that allow you to add or subtract lines of data. The “Pref” radio button indicates which is the preferred line—although any of the words found in the Title fields, for example, would be found in a search for Title keywords. The fields that appear with a small chain link next to them are fields that are linked together; typically, two or more fields will be combined into one larger display field, as seen in the Components/Parts fields in the Work tab. The chain appears broken when the display field has been altered, to indicate that any further changes to the details fields will not be automatically reflected in the display.

As recommended by CDWA, we have incorporated several different types of data entry fields. There are free text fields, which allow the user to enter text in any format. There are also controlled vocabulary fields, where drop-down lists reveal possible entries for that field. There are also authority fields, which are linked to another entry called an authority that provides embedded information about an agent, a built work, a location, or a generic concept. These authority records may be added, edited, and deleted by either selecting “Add New Item” from a drop-down list, or through the Administer button in the navigation bar.

On the Work tab (Figure 8), you can see that there is a field for Catalog Level, which indicates whether the record captures a single item or a group of items. If you had a large group of working drawings as a part of the archive, it would be much easier to include them in one catalog record and then simply specify what drawings are included within the larger group. This is a functionality that is not available to us in our current collection management system, and this kind of part/whole structure is particularly important for architectural archives, as there are typically many layers of materials produced for any given project. Another improvement over CITI is our ability to easily duplicate records and rapidly populate fields that are identical within a group of records by using the “Save As New” button.

The Creation tab (Figure 9) contains information relating to the creation of the work, including the designer(s), date(s), and commissioner information. The creator fields are linked to an Agent authority that includes more comprehensive information about the creator and other agents with a design object.

The Description tab (Figure 10) includes information pertaining to the appearance (physical and virtual) of the design object. Here we see a couple of traditional cataloging fields that are adapted to including digital-specific information, including file size, resolution, and software media.

The Context tab (Figure 11) allows the cataloger to record what exhibitions, books, and photographs also exist that relate to the work.

The Ownership tab (Figure 12) contains fields like acquisition history, copyright owner(s), collection status, etc. that relate to the ownership and legal status of the work.

The Related Records tab (Figure 13) allows the cataloger to connect related records in parallel relationships as well as hierarchical relationships. In this example, the output drawing being cataloged was actually made up of a digital photograph and a rendered scene from Maya, collaged together in Photoshop. Links to each of these native component files are included in this record. The relational database structure of DAArch allows these relationships to be embedded within the object record itself.

The Add Files tab (Figure 14) is where you "attach" files to a record, and also where information specific to born-digital design data would be kept. When the file is uploaded, a copy is sent to the DSpace repository, where it can be accessed via the record at any time. Some of the data entered here in the Add Files screen, as we will see shortly, are essential in helping plan preservation policy and to track what types of files comprise the database.

The fields in the Cataloging History tab (Figure 15) will be automatically populated each time a user logs in.

With the proper privileges, users will also be able to administer controlled lists and authorities. These authorities help maintain consistency among records and allow changes made to an architect's firm name, for example, to be made to all of the records that reference that particular authority. DAArch includes four different authorities: Built Work, Agent, Place/Location, and Generic Concept.

Information about the built work represented in the files is important to record because, of course, we are cataloging the files, the digital design object—NOT the building itself. But through the authority files, we can embed some of the built work information, including address, building type, related agents, and relevant dates each time we enter a record for that particular building. In many cases, as with competition entries, renovation projects, or urban-scale projects, multiple groupings of object records may relate back to a single built entity. The Built Work Authority (Figure 16) allows the basic information to be captured and attached to the record so that it is searchable and multi-layered.

The Agent Authority (Figure 17) records important life data for individuals, families, corporations, and other entities: birth and death locations and dates, life roles, nationality, etc. Agents can include architects but also related entities like donors, curators, graphic artists—and organizations as well.

The Place Authority (Figure 18) contains information relating to a particular geographic location. The majority of data for this authority is imported directly from the Getty's Thesaurus of Geographic Names (TGN).

The Generic Concept Authority (Figure 19) is used for recording definitions and scope notes for important terms; it contains customized terms as well as vocabulary taken from the Getty's Art and Architecture Thesaurus.

Finally, a search screen (Figure 20) allows the cataloger to search by field for any number of words and numbers. Additional search terms may be added or excluded as filters.

**Step 5: Preserving Data**

Once the files have been cataloged and uploaded into DSpace, we need to make sure the integrity of the data is maintained over time, and that our institution's methods of accessing and preserving these data remains up-to-date. Several strategies for digital preservation exist in the archival community. The ability to employ these strategies depends on the level of resources, commitment, and vigilance an institution is able to provide consistently and effectively over the long term. The Art Institute's approach to digital preservation is closely tied to the initial decision to employ a two-tiered collection.

The first, most basic level of preservation that must be maintained is bitstream preservation, the maintenance of the individual bits that comprise a digital file. All files, both native and output, will be preserved at a bitstream level. This task is accomplished by DSpace, our back-end repository. When a file is uploaded using DAArch, DSpace generates and stores a checksum that can be used to verify the integrity of the stored bitstream over time. Periodically, we can initiate a comparison of current and old checksums, to ensure no data has been corrupted or lost.

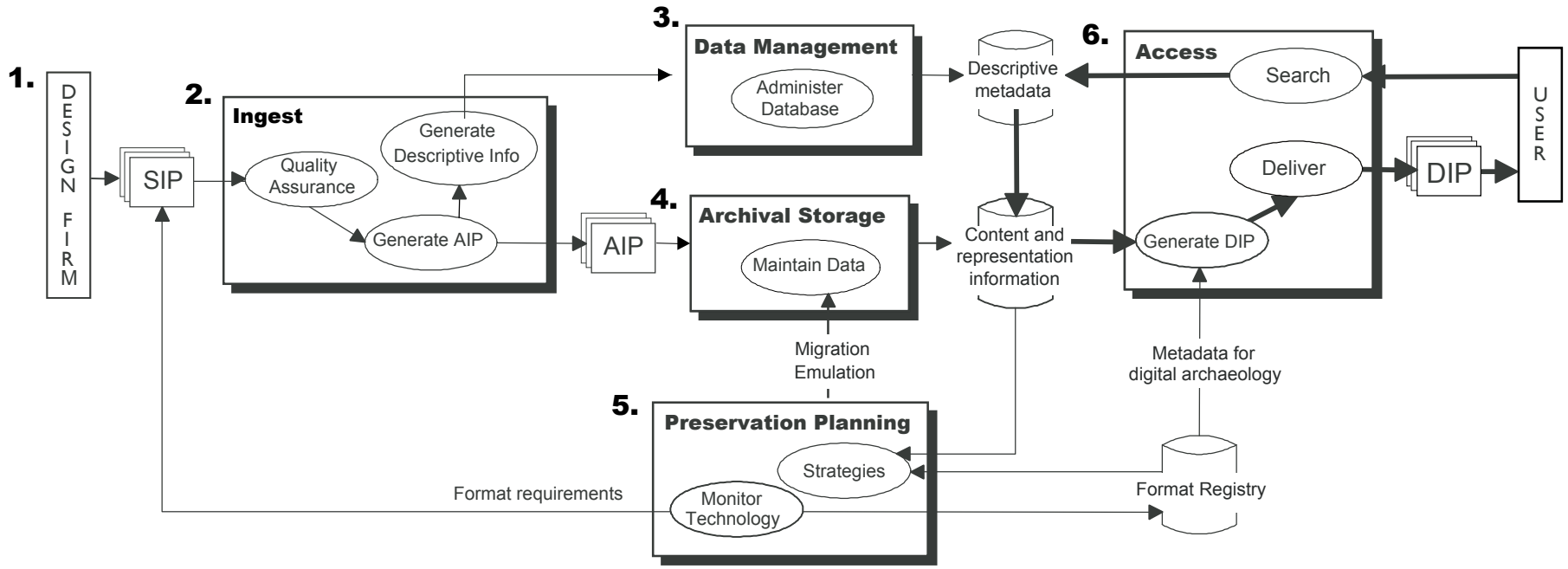
Functional preservation is a slightly different concept: It ensures that data are actually usable (or in this case, viewable) in the same way they were when they was created. Functional preservation can be achieved in three ways: migration, emulation, and translation. Migration involves migrating digital files to new versions of software as they become available. This would mean purchasing upgrades for each type of file in the archive—not a very economical solution. Emulation involves creating a software environment in which the native files behave like they did in the original software. This process is extremely resource-intensive, involving high-level programming and development. Translation, on the other hand, provides a solution for the output files we wish to preserve over the long term. As we saw in the discussion of the two-tiered collection, translating native data into pdf and tiff formats allows us to capture most, if not all, of the features of the digital output.

Making informed preservation decisions is at the heart of any digital preservation project. The last preservation strategy is establishing a Preservation Policy Committee that periodically reviews and adjusts preservation techniques as the technology and practice of architecture change. This committee consists of department staff and practitioners as well as staff from the registrar's office, legal, conservation, information services, and the library. The Preservation Policy Committee may, for example, see that tiff is being supplanted by another, more reliable format. Using DSpace's capability to produce reports that summarize the contents of the database, they might decide to migrate all tiff files, *en masse*, to the new format.

**Step 6: Accessing Data**

For logistical reasons, the pilot project does not include the Access module. Of course, there is still access for end users within the institution—but public access via the web is a step beyond what we wish to accomplish, at least at this point. Our goal is to integrate all of our collection data into one system, to allow each department in the museum to take advantage of DSpace's data management functionality, and to give the public to access our encyclopedic collection online via our web portal.

All of these tasks I've talked about take a lot of resources: time, money, and staff. But they are essential if we want to effectively manage digital assets in a museum or archival setting—to preserve digital materials for the short, if not long term. In the end, I hope that the completion of a successful pilot project will help ours and other institutions develop digital preservation policies and initiate their own digital archives.




SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package

**Steps in the Process:**

1. Preparing
2. Collecting and Processing
3. Cataloging
4. Storing
5. Preserving
6. Accessing

**Figure 1.**

Archiving Workflow Based on Open Archival Information System (OAIS) Reference Model

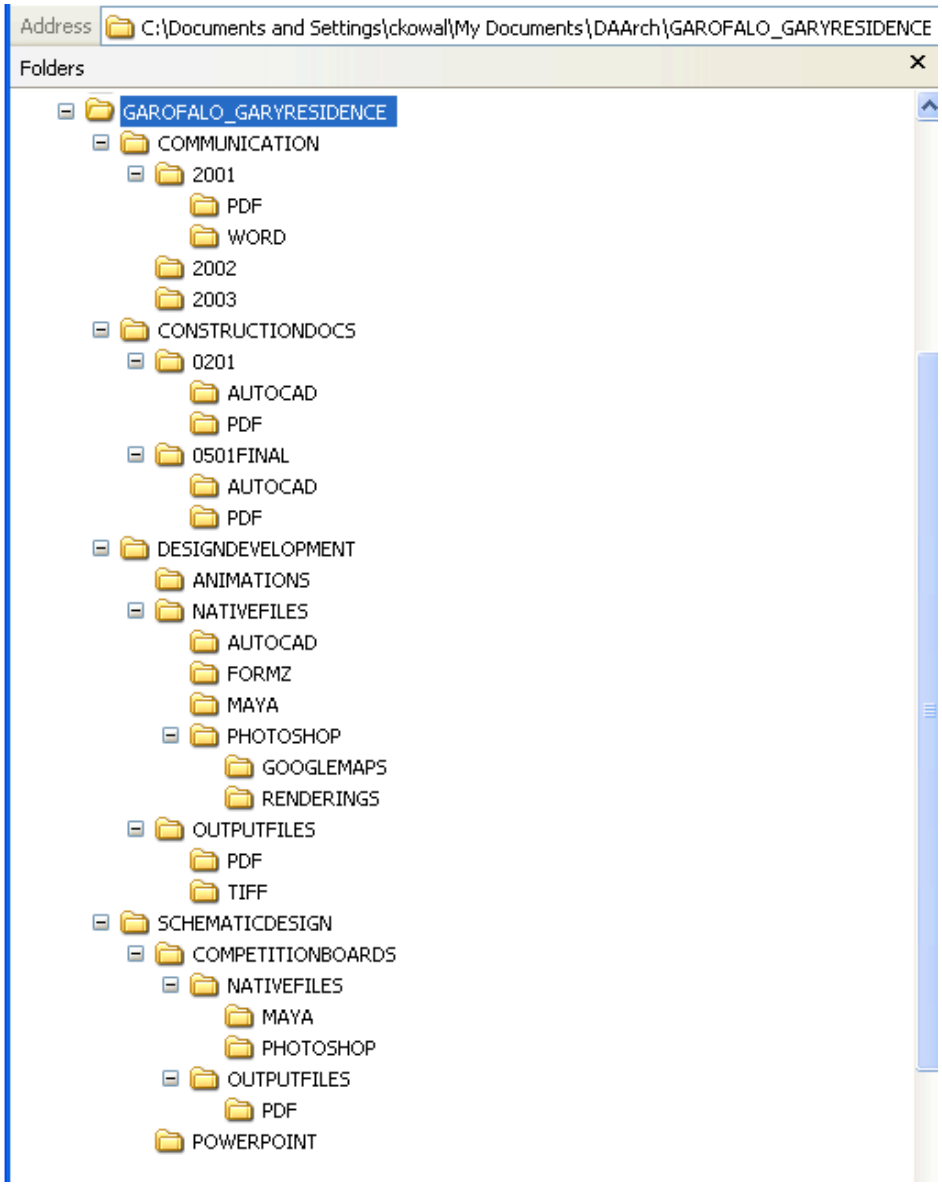



**Figure 2.**

Typical File Structure for an Architectural Project

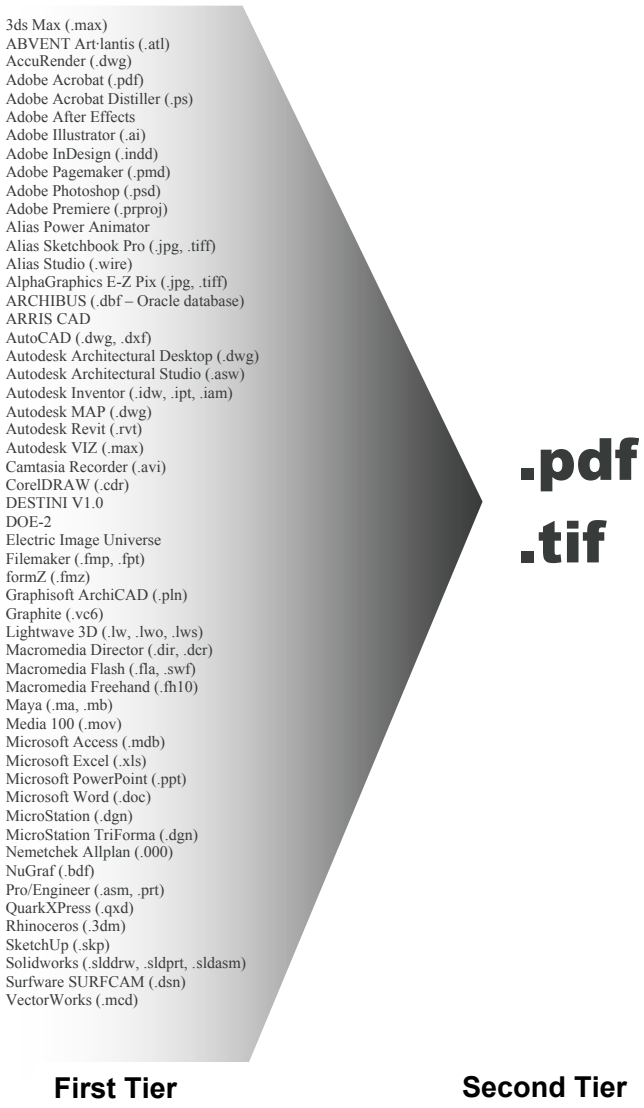



**Figure 3.**

Two-Tiered Approach to Digital Archiving

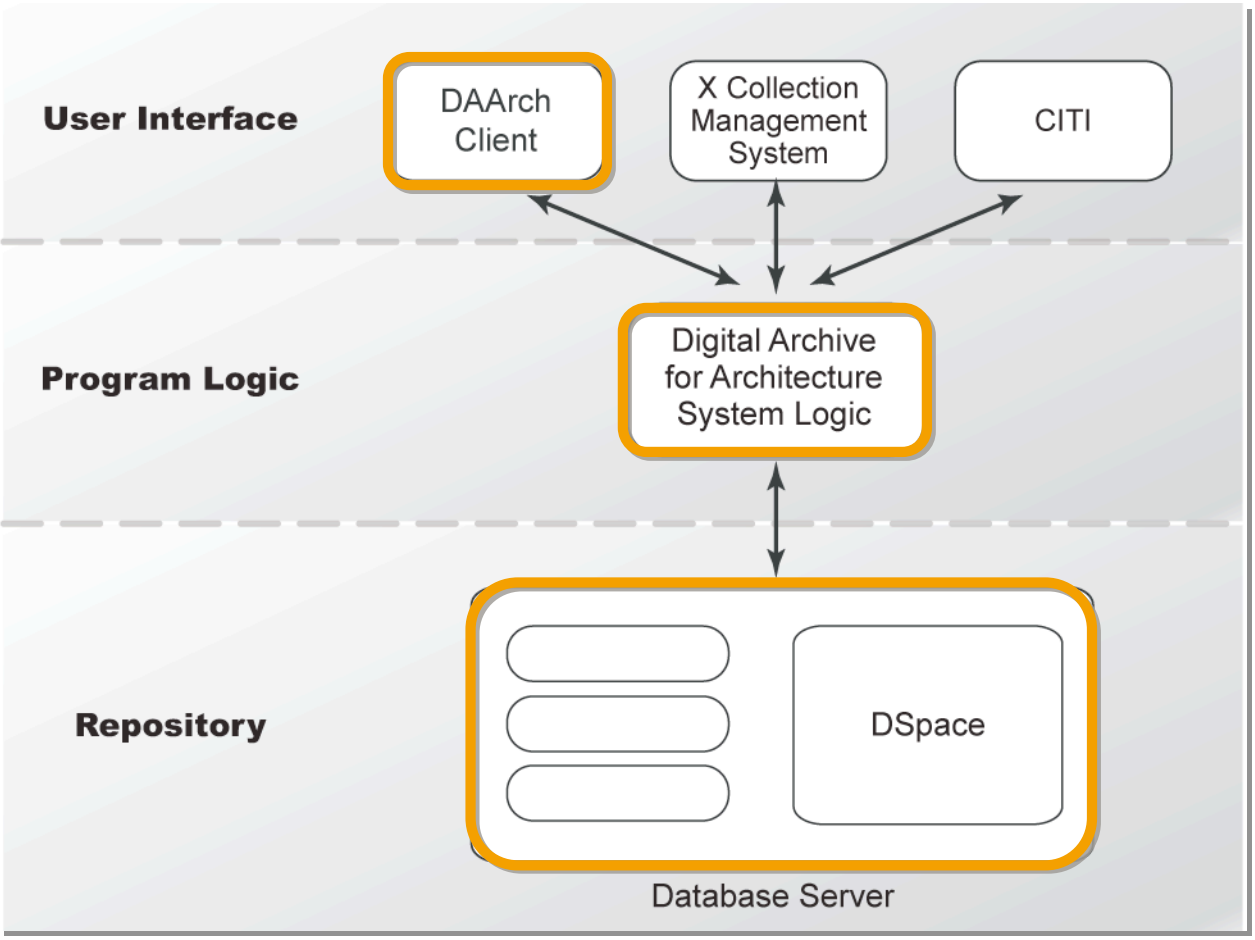



**Figure 4.**

DAArch System Architecture

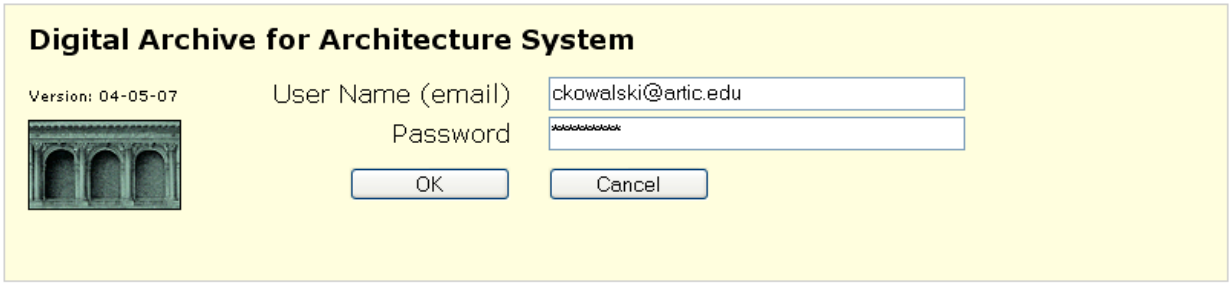



**Figure 5.**

DAArch Login Screen

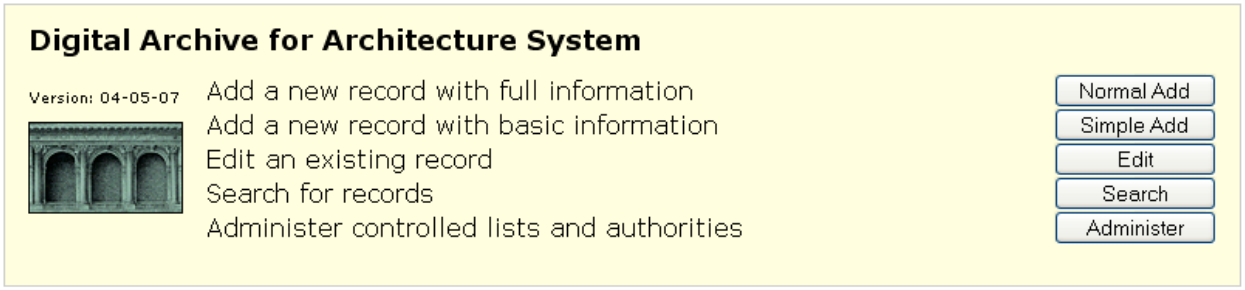



**Figure 6.**

DAArch Home Screen

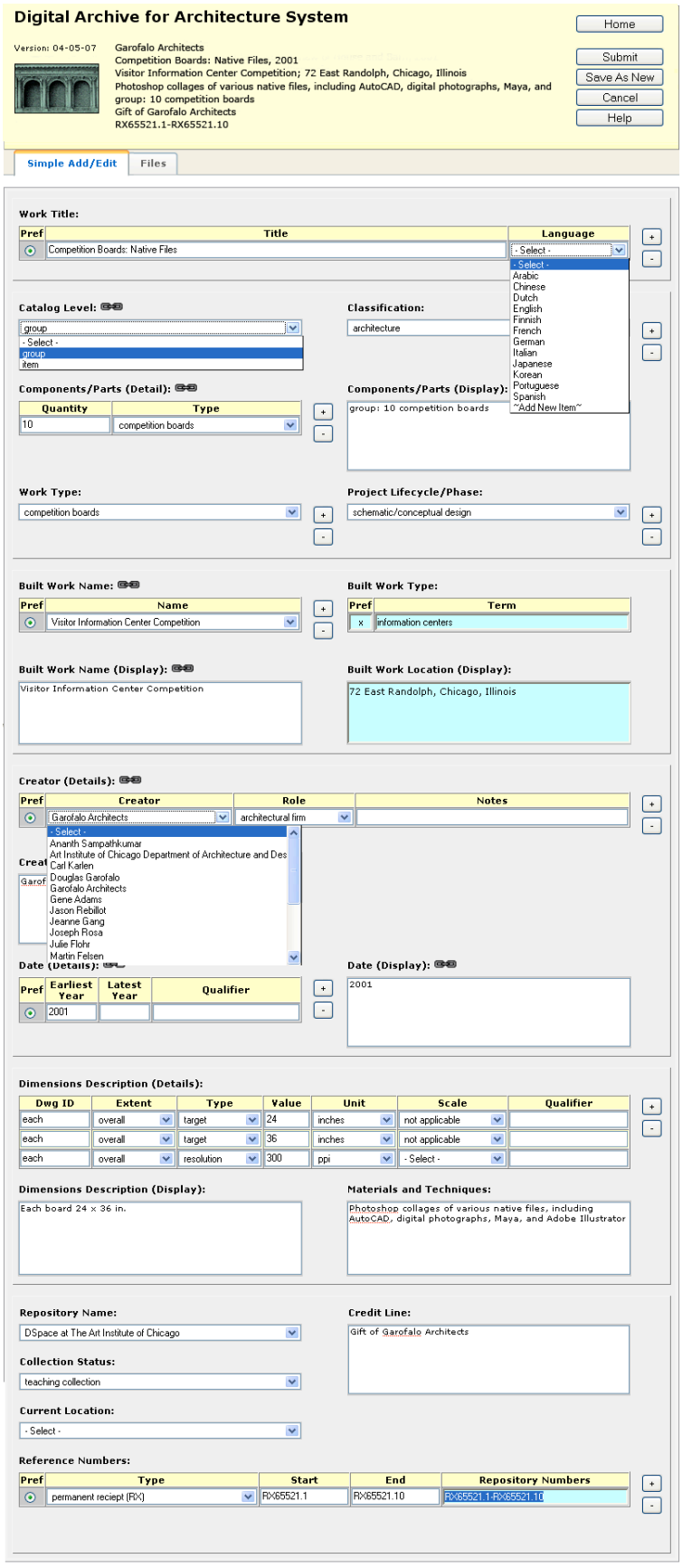



**Figure 7.**

Simple Add

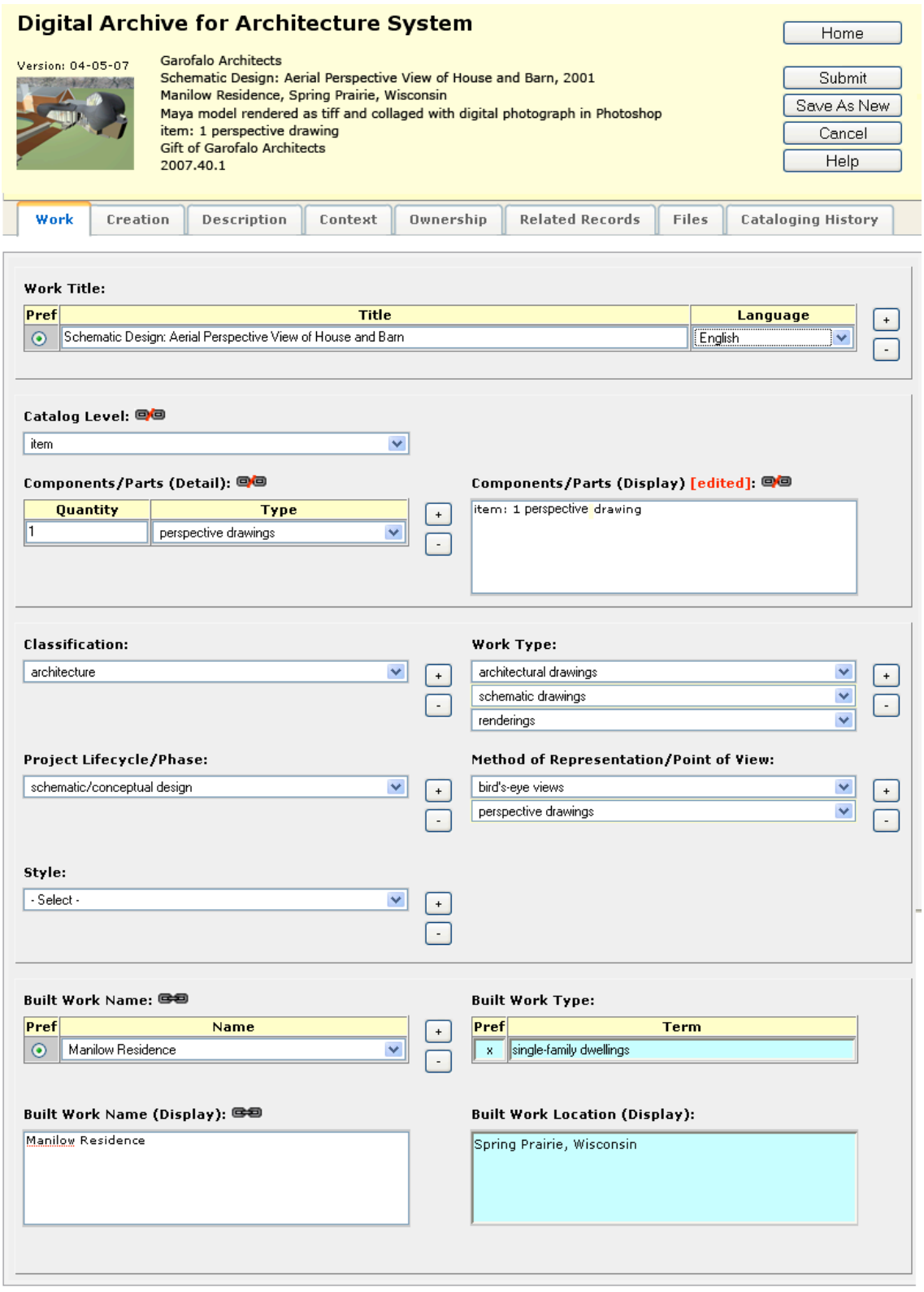



**Figure 8.**

Normal Add: Work Tab

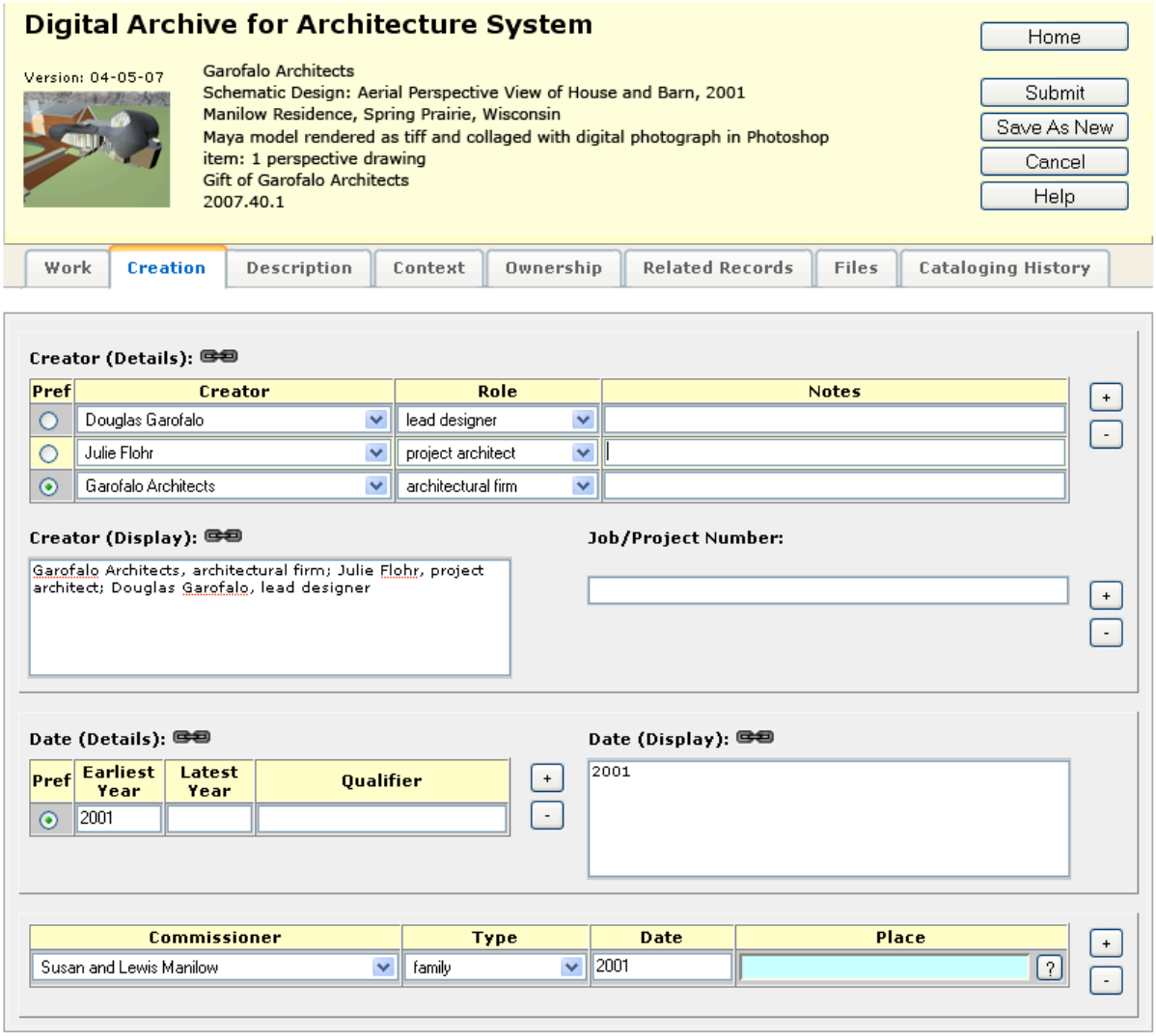



**Figure 9.**

Normal Add: Creation Tab

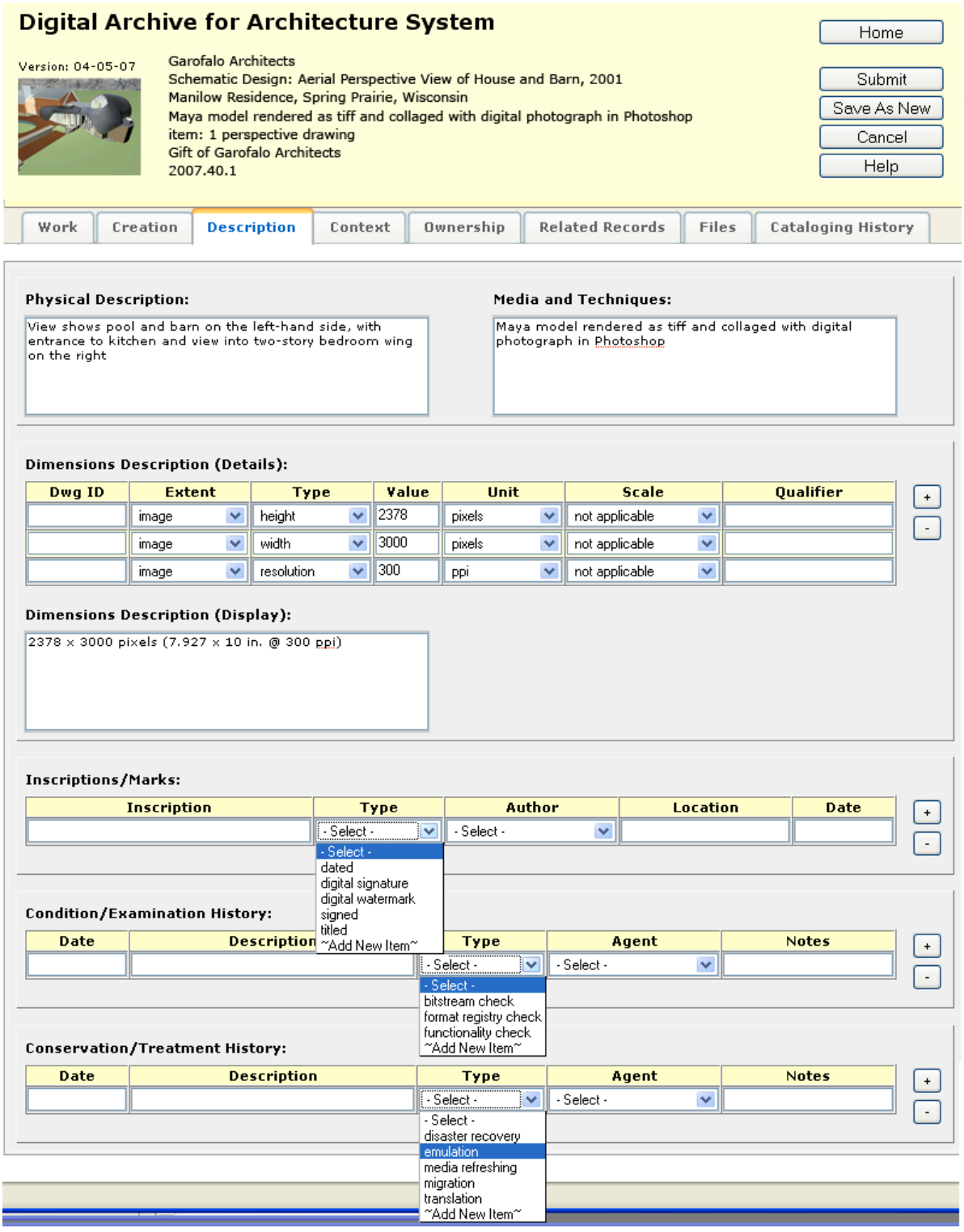



**Figure 10.**

Normal Add: Description Tab

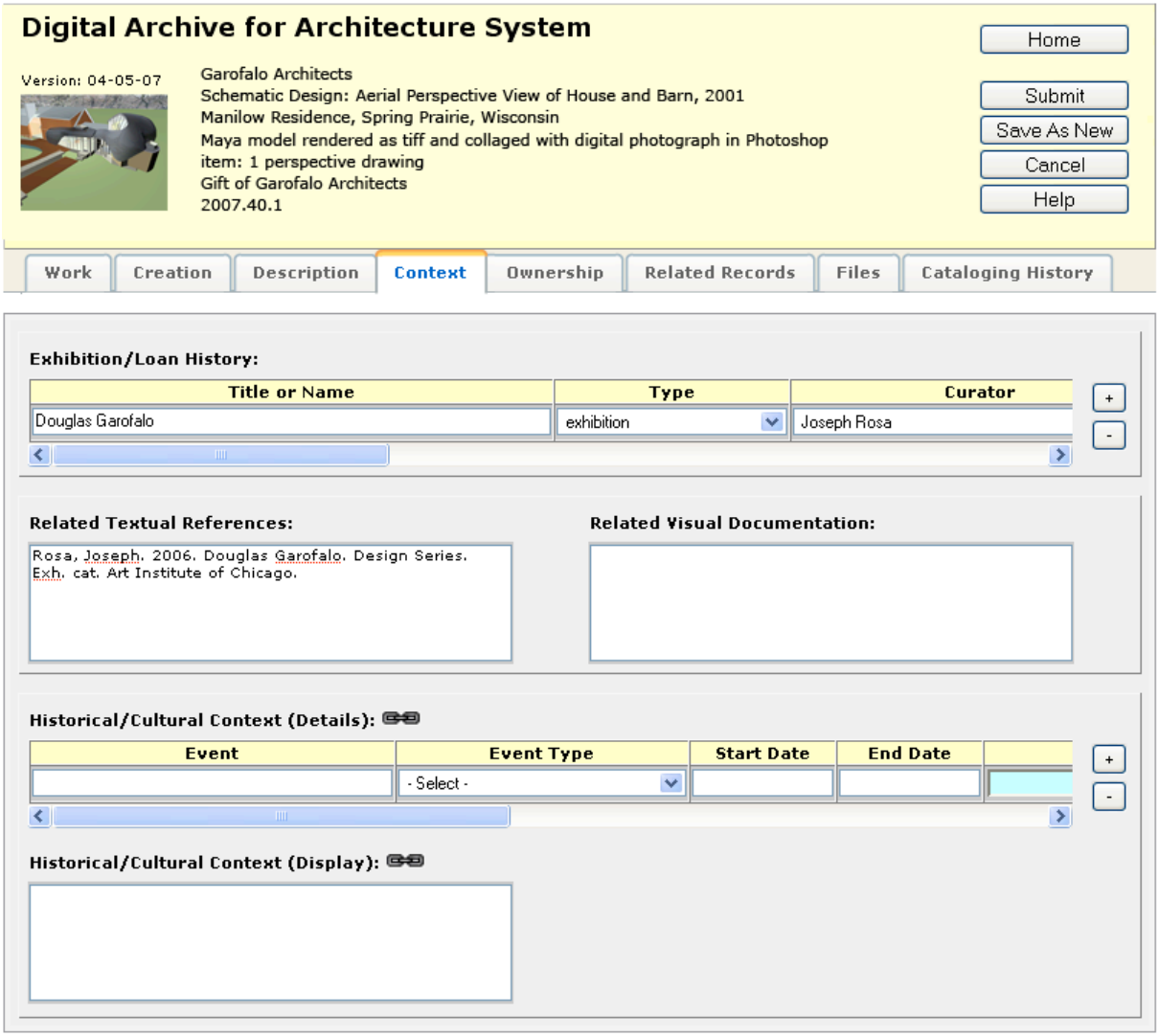



**Figure 11.**

Normal Add: Context Tab

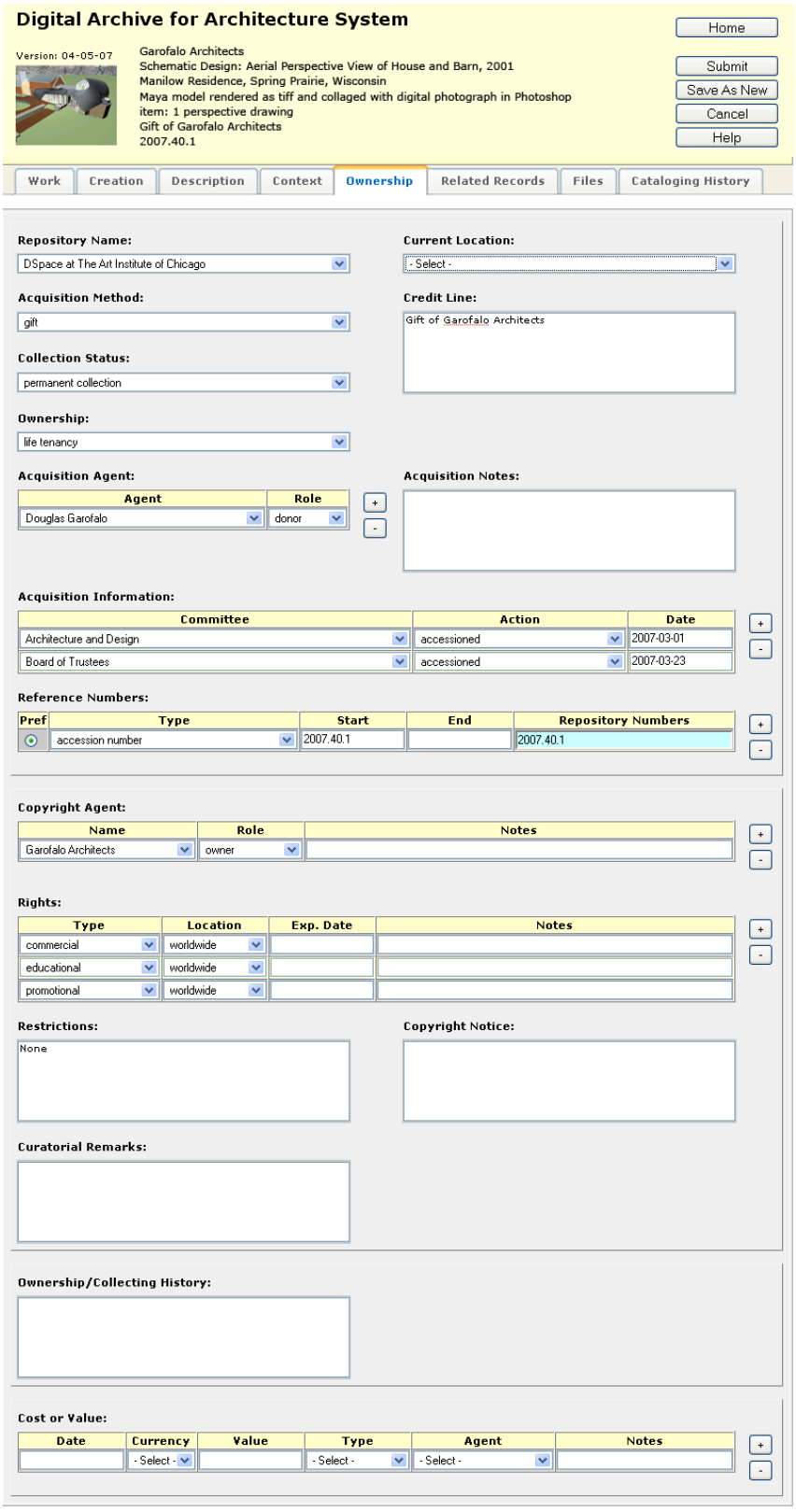



**Figure 12.**

Normal Add: Ownership Tab


**Figure 13.**

Normal Add: Related Records Tab

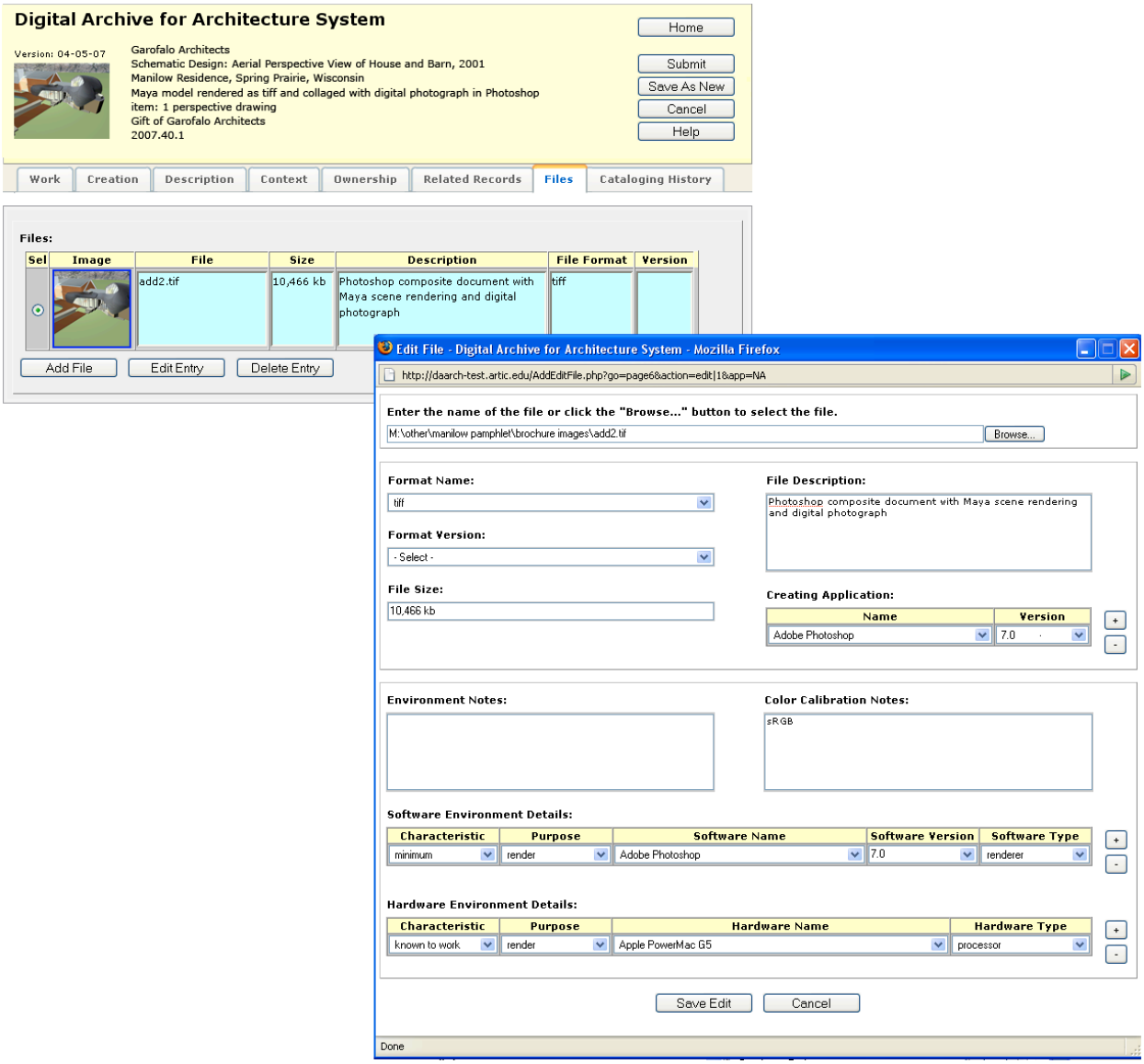



**Figure 14.**

Normal Add: Files/Add Files Tab

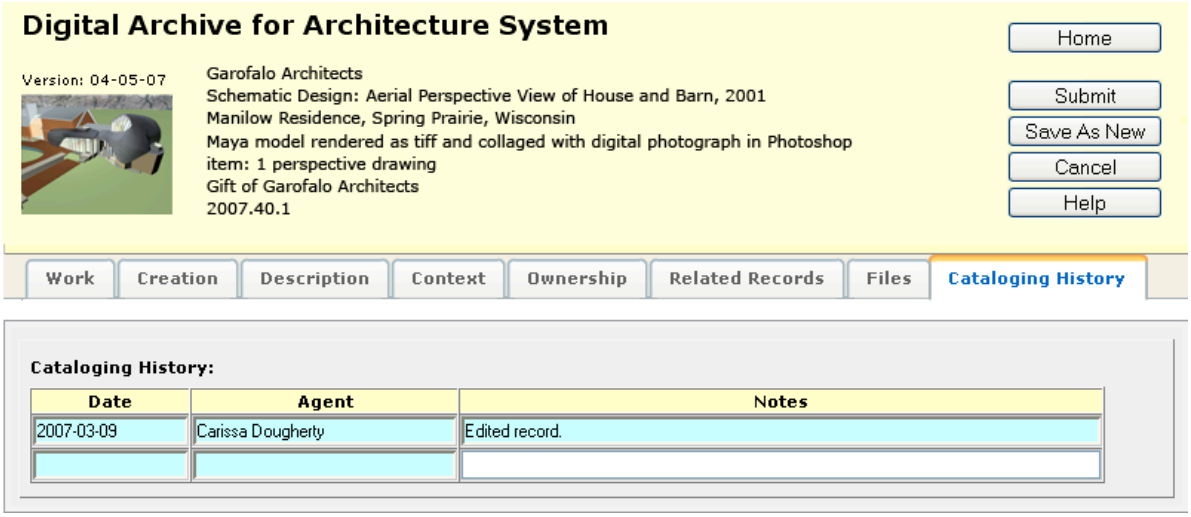



**Figure 15.**

Normal Add: Cataloging History Tab

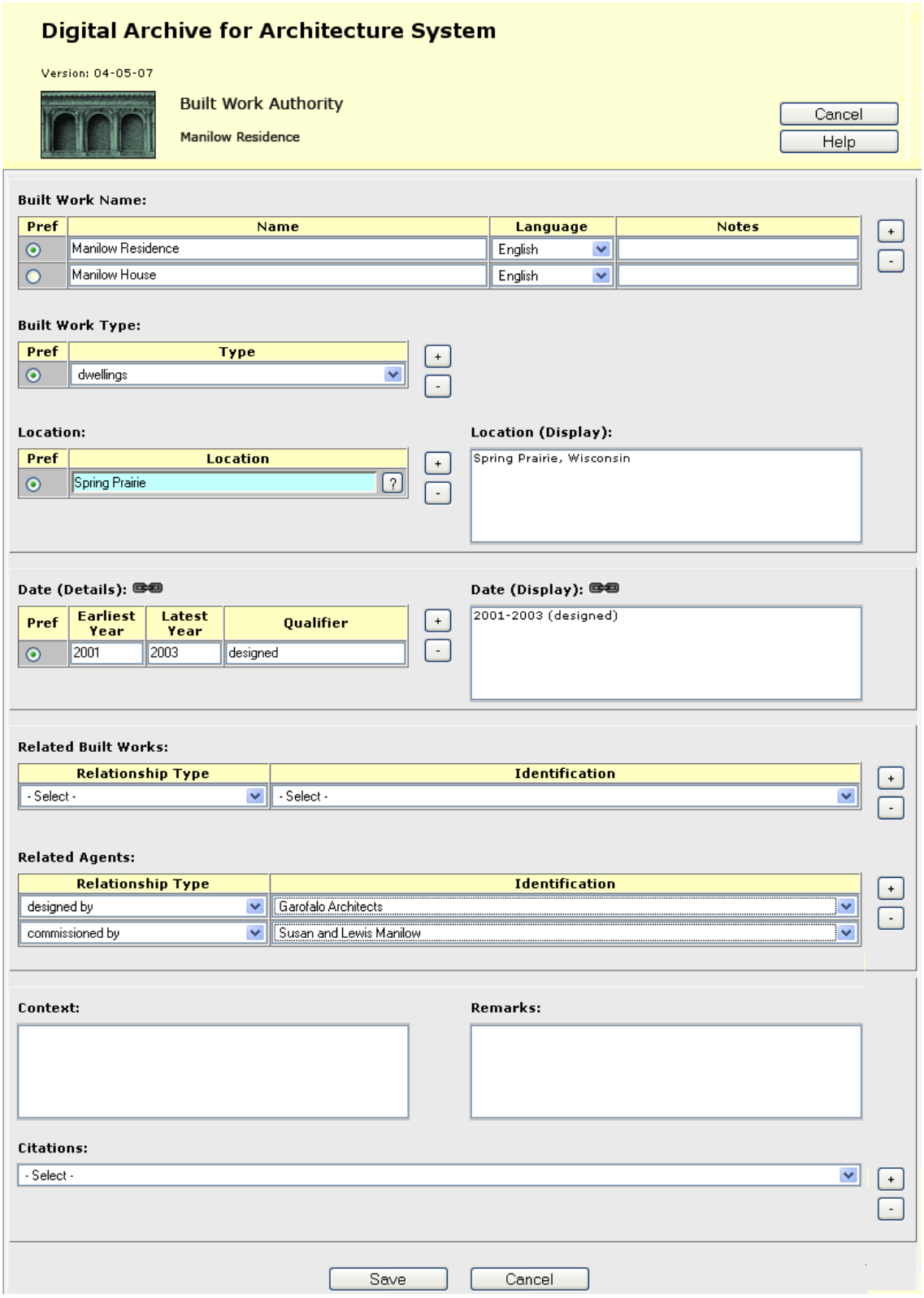



**Figure 16.**

Built Work Authority

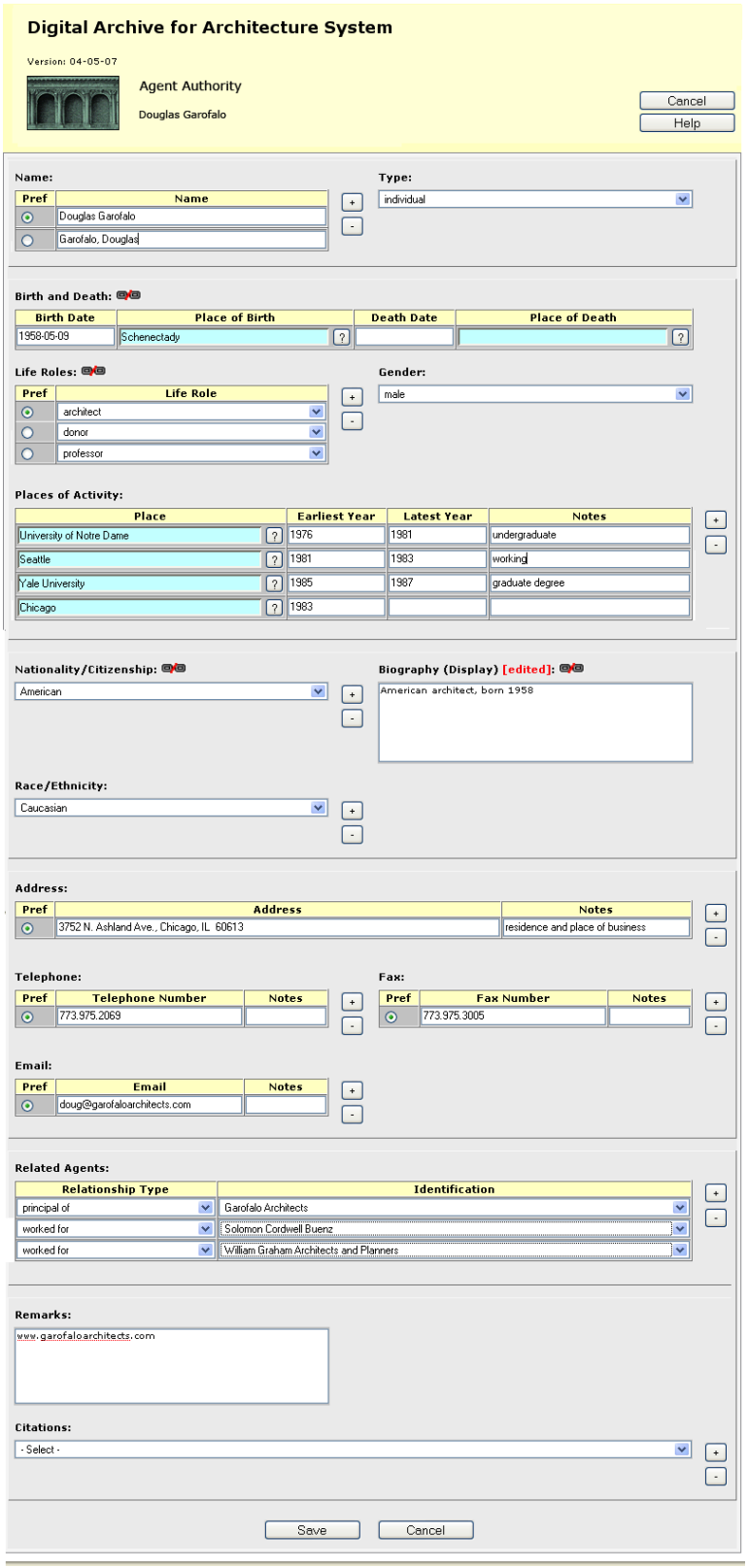



**Figure 17.**

Agent Authority

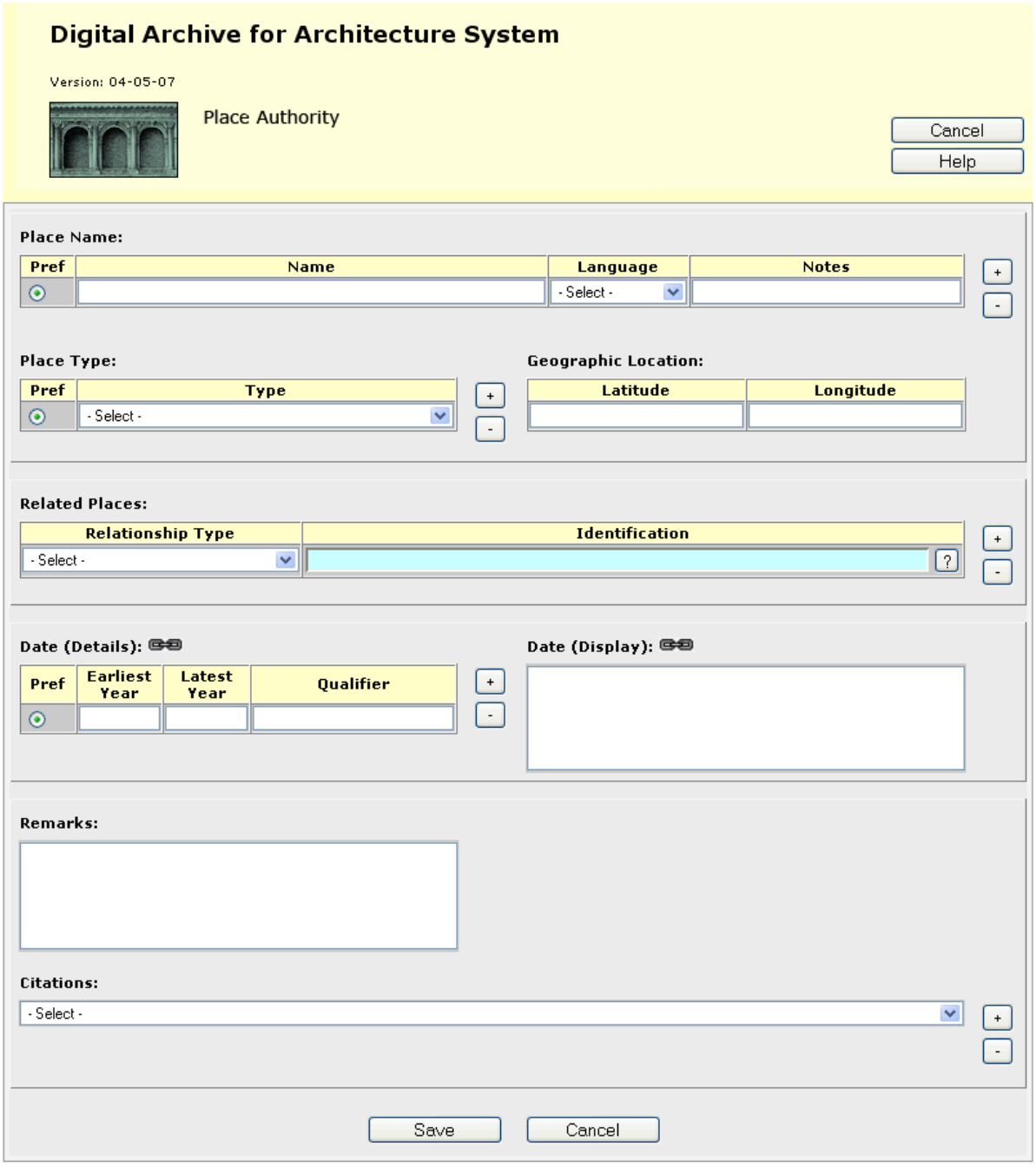



**Figure 18.**

Place Authority

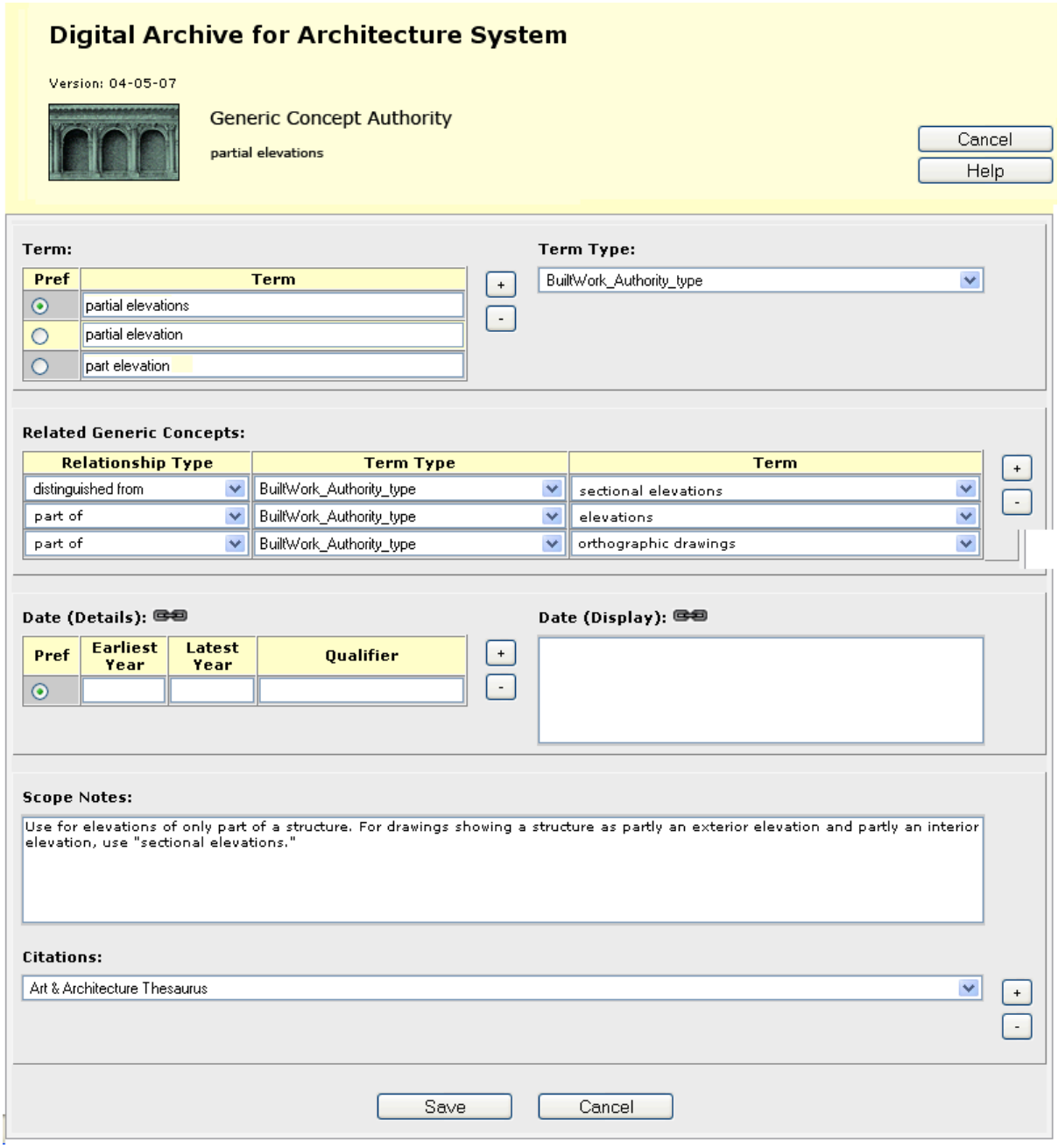



**Figure 19.**

Generic Concept Authority

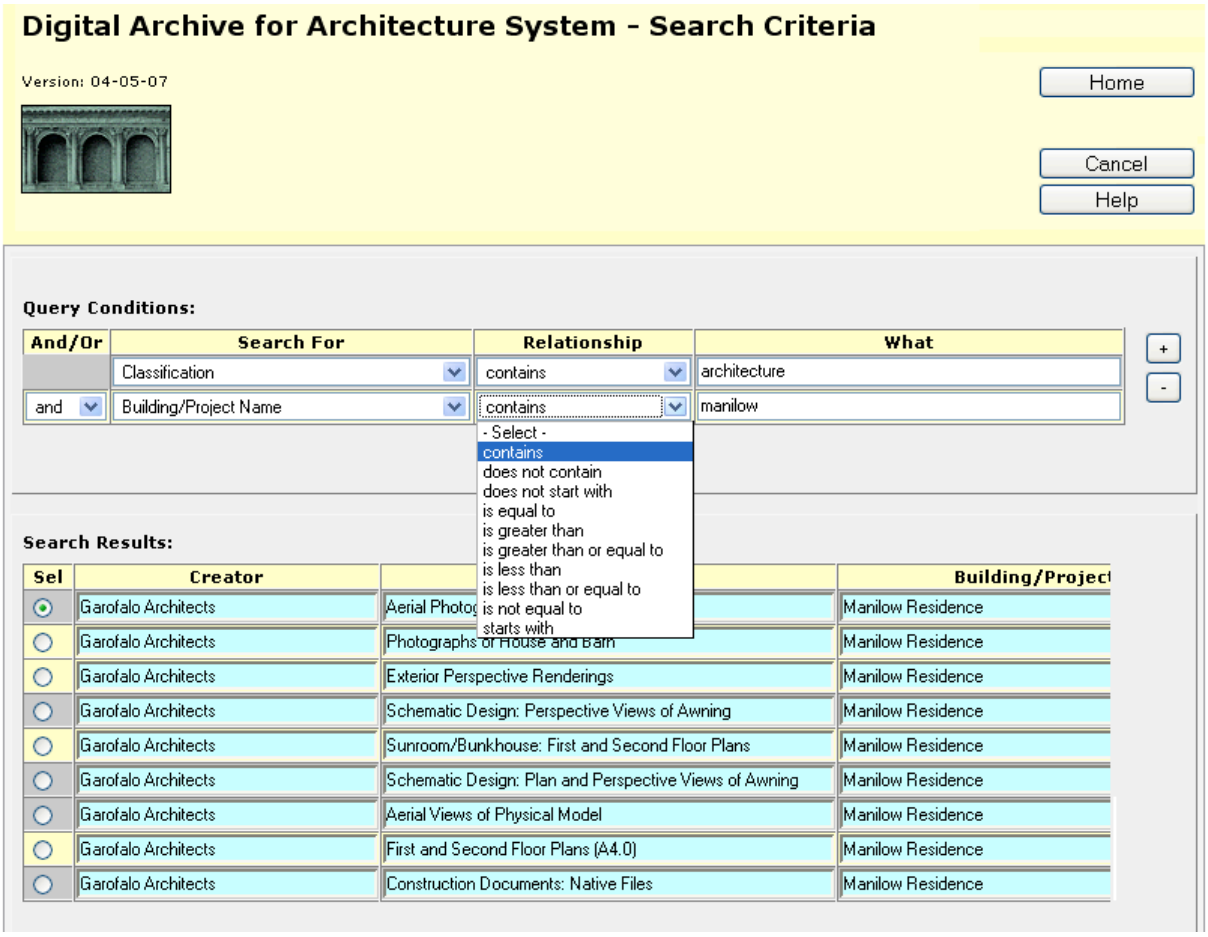



**Figure 20.**

Search Screen

Carissa Kowalski Dougherty is the Digital Archive Coordinator for the Department of Architecture and Design at the Art Institute of Chicago. In the last year and a half, she has worked with software developers and consultants to develop the Digital Archive for Architecture (DAArch) system that will allow the Department to collect and preserve digital design data. The pilot project for implementing this system, which concludes in April 2007, represents a significant step toward maintaining architectural records in the rapidly changing environment of new media and software. Prior to working with DAArch, she served as a curatorial assistant in the Department, helping to organize four major architectural exhibitions.

Carissa has a master's degree in architectural history from the University of Illinois at Chicago and a bachelor's degree in journalism from the University of Wisconsin–Madison. Her research interests include design history, architectural technology, and contemporary theory. In addition to writing several articles in journals and exhibition catalogs, Carissa has also contributed to more general-interest publications, the most recent of which is *Young Designers Americas* (daab 2006).