

IMPLEMENTING A SURNAME STUDY WEBSITE WITH DRUPAL

Degory Valentine¹, Denise Mortorff² and Christophe Giraud-Carrier¹

¹Brigham Young University

²KendallAncestry.com

ABSTRACT

Accessibility of research data in surname-study websites is limited by the technical knowledge of the websites' administrators. We describe the implementation of a surname-study website that leverages Drupal and a powerful taxonomy to make multi-faceted data access and research tractable for personal family historians. The proposed implementation may be packaged into a Drupal module and thus easily reused.

INTRODUCTION

Current on-line genealogical research practices include an area known as surname studies, also referred to as one-name studies. One organization, the Guild of One-Name Studies based in the United Kingdom, has about 7,000 surnames and over 2,000 members [1]. There are also such independent studies worldwide where researchers have created or are establishing research projects to benefit genealogists or family history researchers on-line.

By definition, “a one-name study is a project researching facts about a surname and all the people who have held it, as opposed to a particular pedigree (the ancestors of one person) or descendancy (the descendants of one person or couple)” [1]. Hence, surname study websites are created as a focal point where researchers take it upon themselves to provide a place on the web where other researchers seeking information about the same or a similar name will gain a foundation in what is known of the surname, including archives of relevant records, lineages, genetics, study reports and forums for information exchange.

Interestingly, the primary users of genealogy software are personal family historians, as well as professional and certified genealogists, but not web designers [2]. Yet, the person sponsoring the surname usually serves as the website administrator, 1) deciding the parameters (scope and approach to studying the surname) and format of the website, as well as the technological mechanisms they will employ for fostering research in the surname; 2) determining the size/space and other requirements; 3) selecting the server that can support these; 4) identifying and procuring software for forums; and 5) determining and organizing how the information is delivered. As a result, websites are designed only to the extent of the abilities or the resources of the web administrator or what they are able to garnish. It follows that most of the websites are text-based and never graduate to a higher (e.g., database) level since few family historians or genealogists have the technological skills needed to produce an advanced website.

Searching for information in these studies has been traditionally accomplished with text-based search devices. In particular, single elements are usually accessible, much like one can find

through a browser search, but it is uncharacteristic to find information accessible in which different elements specific to a researcher's interests can be searched in combination. Therefore, the present approaches are cumbersome, at best, and require a lot of reading and sifting through information to find what one may be able to identify and utilize in furthering their research. As an example, in an advanced on-line surname study that focuses on the COMBS surname and its variants [3], a search for the term ARCHDALE returns 730 matches. Since this is text-based, a researcher is compelled to search through as many of the report annotations as possible, until something that sounds like it fits with the application the researcher is seeking is found.

To bring the power of sophisticated indexing, searching, and storage within reach of the typical surname-study website administrator and its users, we have built on the existing strengths of the Kendall Ancestry website, leveraging the power and flexibility of Drupal, an open-source content management system (CMS) for the web [4]. Although illustrated on one specific web site, our pre-configured module set and taxonomy would allow surname website administrators to manage large amounts of historical record data in an extensible framework without being encumbered by Drupal's complexity.

From a family history standpoint, we note that surname studies have not been emphasized in software programs and other on-line resources as tools in promoting genealogy. Emphasis has typically been on storage devices that assist researchers with the organizing and use of their individual genealogical information. Instead, we hope that this work will also open up thinking about websites designed as research tools fostering individual genealogical research. Essentially, the individual gathering and storing personal genealogical information can benefit from the ease of accessibility of records and other information found at websites included in surname or one-name studies.

KENDALL ANCESTRY

The Kendall Ancestry website was designed to answer several of the shortfalls of traditional surname study websites. In particular, it was an essential requirement that record abstracts and reports be defined with relevant facets (to ensure maintainability), and that they subsequently be searchable using these facets rather than text alone (to ensure accessibility).

Records are stored in a *Records Repository* where researchers find a *Record Search* page where the integrated identification of records takes place. Inputs yield a *Search Results* page similar to that utilized in major archives from which specific records of interest to the user can be accessed in the form of a *Record Description*. The identification of records pertinent to researcher interests is further enhanced by the ability of the user to reorder by ID#, Date, Surname, Locality, etc. building in opportunities to narrow down choices that could not otherwise be accomplished readily with text-based searches.

One major advantage of the website is its underlying database, which indeed allows users to select information, singly or in combination, by a specific name, time frame, geographic location, research topic, specific record type, and by record ID#. The search can narrow down possibilities and more readily access information for the researcher. It can also demonstrate what areas contain and do not contain information in the database. For example, if a search for a

specific type of record in a specific place in time returns no results, the user immediately knows that the database does not yet have information within the search parameters without having to sift thru mounds of material. Human errors often introduced by such sifting (e.g., skipping over or missing information) are also less likely with a database. Finally, the database approach makes text-based searches within specific fields possible.

Another advantage at Kendall Ancestry is the design of the *Record Description*, which includes features that enable the researcher to link to research reports within the website, link to scanned images, and locate related resources. Citations and sources are adjunct features of the *Record Description*. They are located at the base of record abstracts and reports, and allow researchers to link directly to related abstracts in the repository, or other resources, entirely bypassing the search page. For example, a researcher would be on a page for Bedfordshire England Probate Records and the discussion would be about the records found in a particular court. A court probate record that had been abstracted and included in the Records Repository will have an ID#. This ID# could be added to the information for a person to link to the abstract directly and appear at the base of a report page with some of the citation information and the ID# to link back to the abstract in the Records Repository.

Our surname-study website implementation has unique strengths, all of which leverage the novel and unique taxonomy developed for Kendall Ancestry. As an accomplished genealogist, the website's designer has organized geographical classifications, stratified by political time period, into a hierarchical *Locality* vocabulary. This vocabulary allows for classification of genealogical information by political regions as they have changed over time. Similarly, a second *Record Type* vocabulary categorizes types of historical records into topical groups. Combining these features enables researchers to identify record groups and specific record types classified within jurisdictions. This approach also allows researchers to more precisely match search parameters to best fit their unique research needs.

LEVERAGING DRUPAL

In going to a content management system, we are able to retain and enhance all of the valuable features of the Kendall Ancestry website, while also creating a system that is sharable and may serve as a model for the rest of the genealogical community.

One of our premises was that in order for the system to be attractive to the community, it had to be free. Hence, we restricted our attention to open source content management systems. After analyzing the main competitors in that space, we settled on Drupal as we felt that:

1. Drupal has the most power and flexibility,
2. Drupal has the best community support, and
3. Drupal seems the most likely to outlast its competitors.

We recognize that this assessment may be somewhat biased by our own experience, and that there may be differences of opinion as to which CMS tool is best. However, we note that it is not so much the tool itself that is at stake here, but rather the concept of automated content management. A result similar to the one presented here could be obtained with another CMS tool.

Our Drupal implementation focuses primarily on two features: the *Record Repository* and the *Research Reports*. The *Record Repository* is an abstraction of record creation and searching. When record descriptions are added to the repository, they are assigned a term from both the *Record Type* vocabulary (see Figure 1) and the *Locality* vocabulary (see Figure 2). Because the type of record is defined by the vocabulary, only a single interface is required for adding a multiplicity of different content.

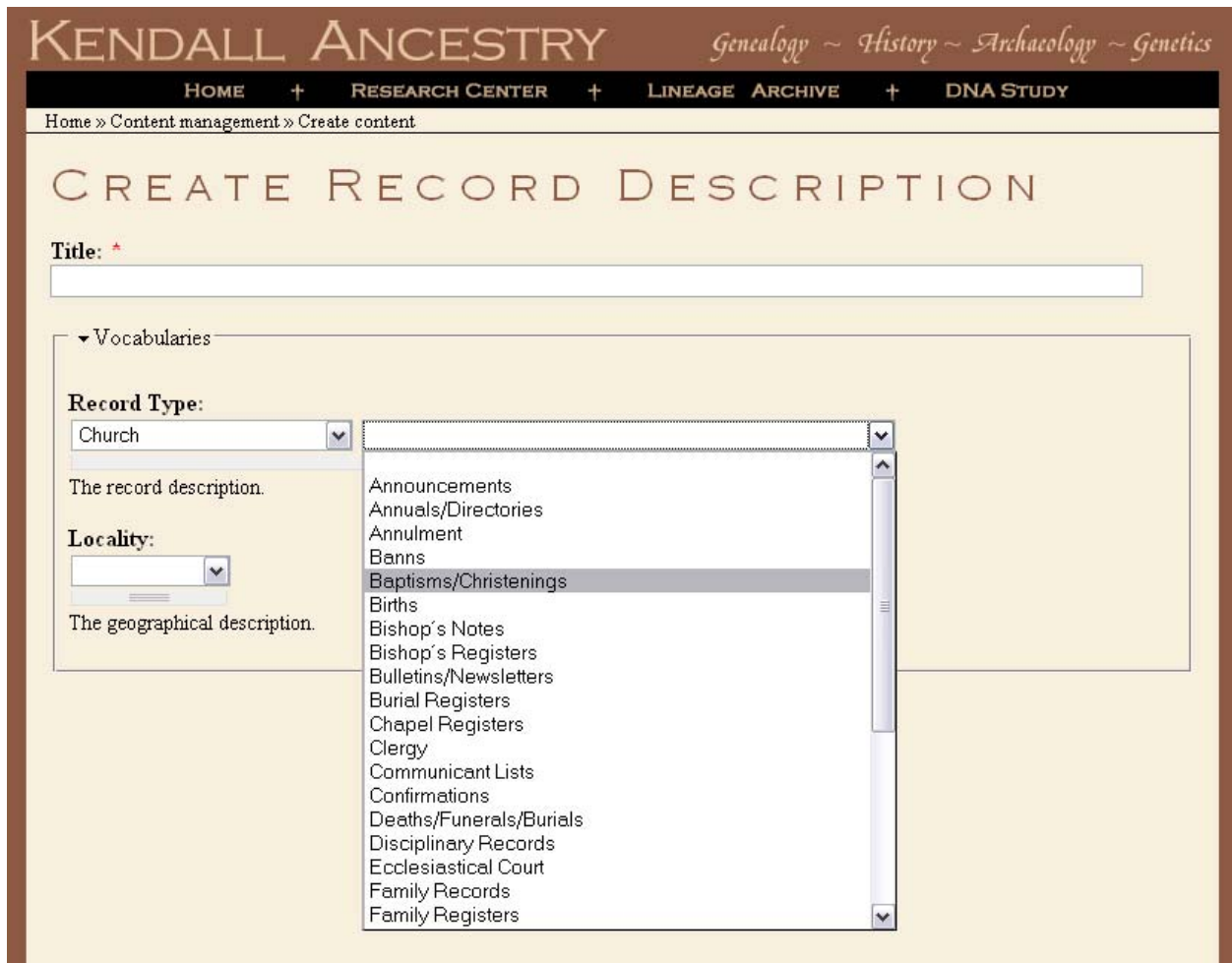


FIGURE 1 - THE RECORD TYPE VOCABULARY

The second salient feature of our implementation is the *Research Reports* section. This section of the site features a browsable hierarchy of research reports organized by *Locality* and *Record Type*. The schema is as follows:

- Country
 - ↳ State/province
 - ↳ Record category



FIGURE 2 - THE LOCALITY VOCABULARY

Since there are 18 record categories defined in the *Record Type* vocabulary, manually maintaining hierarchical links among the research reports as new ones are added would be intractable. Our implementation, however, automatically maintains these links and allows research reports to be added with a singular interface similar to the *Record Repository*, regardless of where the report ends up in the hierarchy. This is accomplished through a set of queries embedded into a single page, taking context from the URL.

To ensure the longevity of our framework, we have used only those modules of Drupal that are most likely to never be outdated. We chose the latest version (as of February 2009) of Drupal, version 6. Three of the four modules used to implement the primary functionality of our site will most likely be incorporated into the core of Drupal, version 7. Table 1 summarizes the primary modules used and their function.

Module Name	Function
Content Construction Kit (CCK)	CCK enables us to add custom fields to our content. This answers our need for common facets across different content-types such as the Record Descriptions and Research Reports.
Views	Views enable us to control how lists of content are presented. The parameters of the lists can be preset, or left for the user to choose. This answers the need to search content using common facets and facilitates easy browsing of content.
Panels	Panels are containers for content layout. They allow us to pass context-sensitive parameters from the URL to many Views on a single page. This answers the need for easy browsing and accessibility.
Taxonomy Redirect	Taxonomy Redirect changes how taxonomy terms are linked within Views. This allows us to have clean-looking, automatically-linked content lists.

TABLE 1 – PRIMARY DRUPAL MODULES AND THEIR FUNCTION

In terms of development, it took an experienced programmer about 3 months to analyze the existing Kendall Ancestry website and to re-implement it using Drupal. That implementation can be packaged into a Drupal module and hence easily re-used. Indeed, tools like Drupal are within the reach of savvy non-expert users who are willing to give them a try. Only the addition of functionality beyond standard Drupal modules would require a programmer's assistance.

MEETING THE COMMUNITY'S NEEDS

While avoiding the (somewhat obvious) pitfalls recently identified for the specific design of one-name study websites [5], our implementation also meets most of the criteria identified for successful genealogy software [2], as follows:

- The software we use should teach us how to research Family History.
 - An important part of learning is understanding the related taxonomy.
 - Users are exposed to the taxonomy through search menus and browsing.
- It should *organize*.
 - The two vocabularies of the taxonomy form the fundamental organization.
- It should *transcribe*.
 - Abstractions of historical records in the *Record Repository* include images, vocabulary terms, and other essential facets.
- It should *cite sources*.
 - Citations are a facet of the record abstractions.
- It should provide *source provenance* (origin).
 - The current location of the actual record is a facet of the record abstractions.

- It should *remember*.
 - As with every content management system, audit trails, logs, and usage statistics are readily available.
- It should *share*.
 - Access privileges can be controlled with fine granularity.
 - The configured modules can be rolled together into a ready-to-use module.

RELATED WORK

There are a few noteworthy projects that approach a solution similar to the one we describe here. Some family historians have implemented their surname-study websites using a variety of content management systems. For example, the North Cheshire Family History Society (dnr.me.uk/ncfhs2) uses Joomla! [6]; the Johnson Family History (johnsonfamilyhistory.org) uses DotNetNuke [7]; the Bunting Society (buntingsociety.org.uk) uses Geeklog [8]; and Think Genealogy (thinkgenealogy.com) uses WordPress [9]. Although these sites are also extensible to varying degrees, they lack the added value of carefully designed module selection and configuration that make our implementation unique.

Other family historians have also designed genealogical software that can plug into CMS frameworks. One significant example is a pedigree-centric genealogy program called The Next Generation (TNG) [10]. It offers powerful pedigree and timeline views of genealogical data. However, TNG cannot support the type of multi-faceted record repository needed for comprehensive, surname research. Drupal can be extended (through its modular framework) to integrate with TNG and administer access privileges to TNG resources [11].

From the perspective of supporting family history, one may argue that there are (at least) three types of things that one may wish to accomplish:

1. Present an individual's personal pedigree(s) with relevant records. If records are provided, they are noted only if they have relevance to the lineage studied or have potential for such relevance. This is the domain of standard software packages such as PAF.
2. Present numerous pedigrees with records for the purpose of linking specific identities of people into specific lineages. This seems to be the approach currently taken by TNG.
3. Present records for researchers to determine the relevance to their own research. The basic premise is that the site is there to help people, but not to do their work; individuals must become experts at tracing their own family history. Lineages are included of course, but they are treated as a record which researchers are expected to use as an informational tool in figuring out their own connection. Website can include lineage studies by lineage experts, but it is not the objective of the administrator to assume expertise in establishing the identities of persons in lineages and records except for their own perhaps.

One-name study websites, such as Kendall Ancestry, are squarely set in the third category. The use of a CMS to underly one-name study websites enhances these sites' ability to deliver on their promises, and to do so in an administrator-friendly way.

CONCLUSION

The Drupal implementation presented in this paper answers the need of surname-study website administrators for a more powerful data access method. It also uses the Locality and Record-Type vocabularies to allow all record descriptions and research reports to be added from a single interface for each. It puts a powerful and easy-to-use website within the technical reach of the personal family historian.

From a user's perspective, the contributing factors that make the CMS/database implementation more useful (i.e., realize its potential) are the general website organization, the nature of the page content and presentation, and the connectedness between website areas. These, and other factors, allow users to better study and coalesce information to see its value in researching family history.

REFERENCES

- [1] The Guild of One-Name Studies. Online at <http://www.one-name.org/>.
- [2] Tucker, M. (2008). 10 Things Genealogy Software Should Do. In *Proceedings of the Eighth Family History Technology Workshop*.
- [3] Combs, Coombs &c. Online at <http://www.combs-families.org/combs/index.html>.
- [4] Drupal. Online at <http://drupal.org/>.
- [5] Whitaker, S. (2008). How Not to Design a Website for your One-Name Study. *Journal of One-Name Studies*, July-September, 20-21.
- [6] Joomla! Online at <http://www.joomla.org/>.
- [7] DotNetNuke. Online at <http://www.dotnetnuke.com/>.
- [8] Geeklog. Online at <http://www.geeklog.net/>.
- [9] WordPress. Online at <http://wordpress.org/>.
- [10] Lythgoe, D. (2009). The Next Generation of Genealogy Sitebuilding. Online at <http://lythgoes.net/genealogy/software.php>.
- [11] Ramos, A. (2009). The Next Generation Integration. Online at <http://drupal.org/project/tngintegrate>, (Accessed 30 Jan 2009).