

FamilySearch Indexing: Providing Easy Access to Your Records

Scott Flinders; FamilySearch; Orem, Utah/USA

Abstract

Searchable indexes and digital browse aids play an essential role in providing easy access to an archive's records. FamilySearch is making tools available to archives to help them index their records. Learn what the FamilySearch indexing system is and how it can help you add digital browse aids to published images, create searchable indexes of records, and link existing records or indexes to images.

FamilySearch has different indexing environments and deployment options available to fit the needs of all archives, regardless of your size or location. You can use the system to either recruit and manage your own volunteer workforce or leverage the 200,000+ volunteers already involved in the indexing program.

Value of Searchable Indexes and Digital Browse Aids

Records archives around the world store and preserve billions of historical records which are extremely valuable to researchers. However, it can be very time-consuming to “find the needle in the haystack” if the researcher has to search page by page, volume by volume through all of the records to find what he or she is looking for. By providing searchable indexes and digital browse aids for the records in their collections, archives provide a valuable service to their patrons. Furthermore, these tools simplify the lives of the archives because patron requests are handled in a much more efficient manner.

For those that may not know, let's start with a quick definition of searchable indexes and digital browse aids.

Indexing is the process of transcribing the key fields in a record or digital image. It is not a complete transcription, but a capture of the key elements of the record so it will be searchable. Key fields may include people's names, place-names, dates of important events, etc. These indexes are then stored in a searchable database. Patrons use a simple form to enter the information they are looking for and are shown all possible matching records.

Digital browse aids are tags assigned to groups of images within a collection—something like chapters in a book. By providing these browse aids, patrons can quickly jump to a smaller set of images within a larger collection.

Why Is FamilySearch Involved?

FamilySearch is the largest genealogy organization in the world. Millions of people use FamilySearch records, resources, and services to learn more about their family history. For over 100 years, FamilySearch and its predecessors have been actively gathering, preserving, and sharing genealogical records worldwide. Patrons may freely access our resources and service online at FamilySearch.org, or through over 4,500 family history centers in 70 countries, including the renowned Family History Library in Salt Lake City, Utah.

FamilySearch is a service provided by The Church of Jesus Christ of Latter-day Saints. Our commitment to helping people connect with their ancestors is rooted in our beliefs. We believe that families are meant to be central to our lives and that family relationships are intended to continue beyond this life. We therefore believe that all family members—those living, those past, and those future—share an enduring bond that reaches across the generations.

We encourage all people to seek out their ancestors and preserve their family histories. Because interest in family history is not limited by culture, ethnicity, or religious faith, we welcome all who want to discover more about their family and their heritage.

More Access to More Records for More People

The FamilySearch Records Access program is designed to significantly increase the number of genealogical records available to people around the world. Through the program, FamilySearch teams up with records custodians and other organizations to gather, preserve, digitize, index, and post the records online. The records can then be shared through the Internet for convenient public access.

The program provides a platform for FamilySearch and records custodians to combine and integrate their efforts. FamilySearch acts as both a facilitator and a participant by leveraging its knowledge, resources, and relationships in the process. The strength of the program is its collaborative approach—an approach that ensures participating organizations can more effectively and economically meet their goals.

In cooperation with legal custodians of records worldwide, FamilySearch has over 200 cameras currently digitizing records in 47 countries.

As part of this digital conversion and capture project, they are currently producing over 60 million images per month or approximately 370,000 rolls of microfilm per year, the equivalent of about 6 million 300-page volumes. Names become the primary focus of the searchable databases; localities and jurisdictions become identifiers (also known as digital browse aids).

FamilySearch indexing is a worldwide, non-profit community effort to harness volunteers to gather, transcribe, and index records of genealogical significance. Using the online indexing system, volunteers from around the world are able to quickly and easily transcribe the records—all from the convenience of their home. The indexes are then posted for the public to use free of charge at FamilySearch.org.

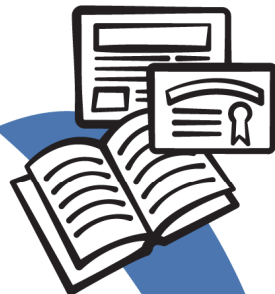
While “crowdsourcing” is an overused buzz word for business these days, it is nevertheless a powerful concept for combining the efforts of the community in a worthwhile activity. FamilySearch has found this to be the case in organizing efforts

FROM ARCHIVE TO ONLINE

INDEXING MAKES IT POSSIBLE

HISTORICAL RECORDS

HANDWRITTEN AND
HARD TO USE



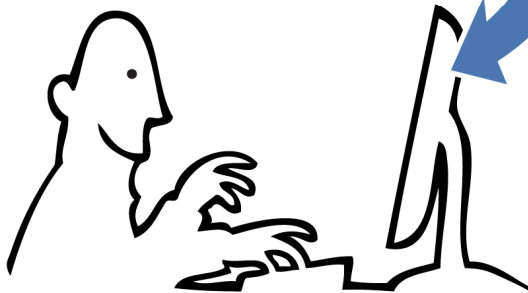
DIGITAL CONVERSION

DIGITAL IMAGES, BUT STILL
NOT SEARCHABLE



INDEXING

VOLUNTEERS TRANSCRIBE
INFORMATION FROM DIGITAL
IMAGES MAKING IT SEARCHABLE



ONLINE ACCESS

INSTANT ACCESS TO YOUR
ANCESTORS – FREE AT FAMILYSEARCH.ORG

2.5 BILLION NAMES AVAILABLE

SEARCH 

around indexing family history records and essentially bringing to life the names and stories of ancestors previously hidden in the masses of paper records in libraries, archives, public institutions, etc.

Online crowds of FamilySearch indexing volunteers, from all walks of life and countries around the world, help index millions of digital images of historic records in seven languages for free access by the worldwide genealogy community on FamilySearch.org. Through the efforts of these amazing volunteers (more than 160,000 volunteers participated in 2011), over 2.5 billion searchable names from over 535 record collections can be accessed online for free by genealogists in the Historical Records section of FamilySearch.org. Thousands of new volunteers continue to join the FamilySearch indexing initiative each month, so the numbers of accessible, free genealogy records will only continue to grow.

We believe that the crowdsourcing aspects of indexing records will grow in the coming years to add functionality and new ways to make the records more usable for researchers and the public alike.

Introducing a New Indexing System

The current FamilySearch indexing system has done a wonderful job of supporting a growing workforce of active volunteers and allowing us to index hundreds of millions of records in 11 languages each year. However, the system is nearing its functional capacity, and it is time for us to roll out a bigger, more robust system with a new architecture and features better suited to handle the continuing growth in users and projects.

One of the goals we have for the new indexing system is to broaden the crowdsourcing efforts within the larger community of records archives and genealogy organizations.

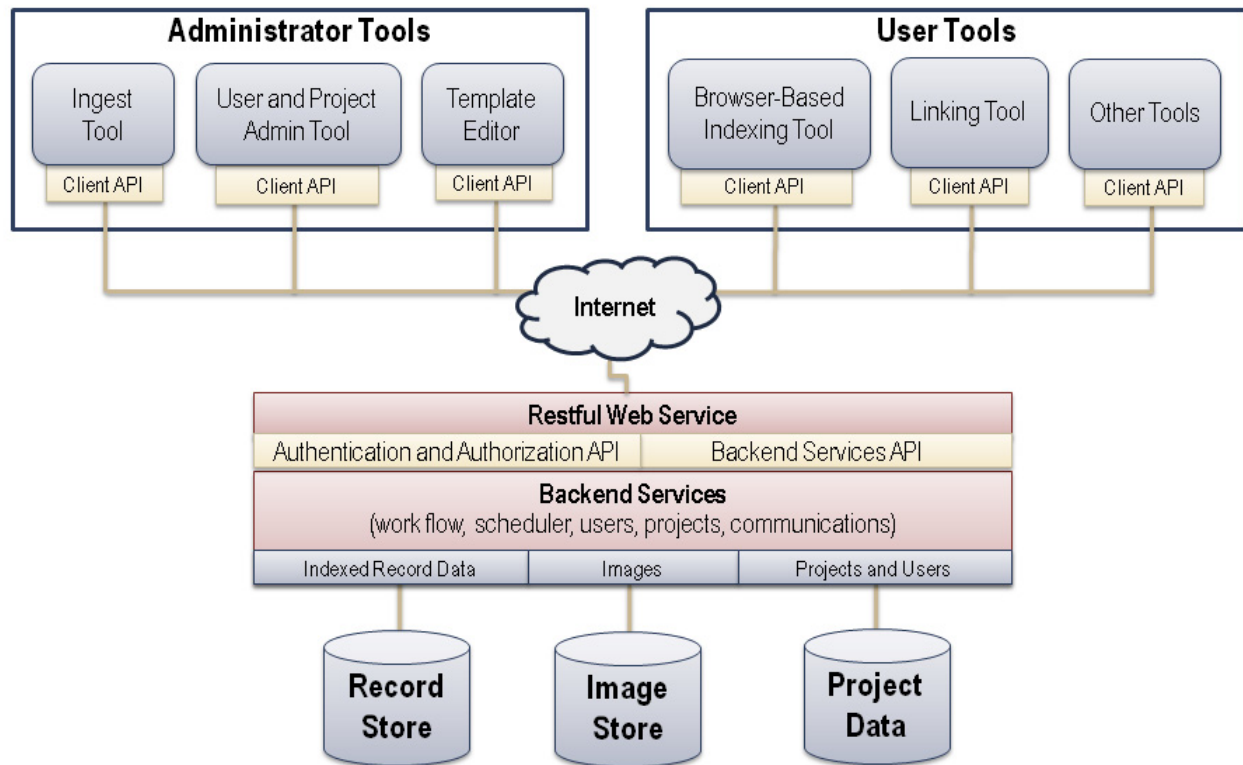
The new system was designed to support multiple large projects similar to those currently processed by our system today as well as smaller projects of local interest created and managed by independent organizations. Some of the key components of the new system include

- Cloud Infrastructure
 - Horizontal and vertical scalability
 - Support for multiple instances and “private label” hosting
- Simple Project Setup
 - Societies and affiliates set up their own projects
 - Indexing on the fly
 - FamilySearch administrative bottleneck reduced
- Clean Code
 - Better testability and unit test coverage
 - Better design for new features (highlight areas, linking, group management, digital browse aids, etc.)
 - Maintainability
- Formal API
 - Better load and integration testing
 - 3rd Party application support and integration

System Architecture

The enormous size and scope of this worldwide effort has driven the architecture and design of the next generation of FamilySearch indexing and related technologies. The system architecture shown below is designed to meet these demands.

FamilySearch Indexing System Architecture



FamilySearch anticipates large numbers of these systems running throughout the world, servicing the needs of all indexers.

FamilySearch has designed the new indexing system to take advantage of the Internet, while at the same time allowing for offline work where Internet access is limited or not allowed. Further, the restful web service architecture anticipates the future desire of individuals to creatively find and build new tools to index records, such as mobile devices or games. Back-end services to images and other indexing-related information will be provided to help solve the difficult challenges the indexing infrastructure requires.

Full Suite of Indexing Tools

The FamilySearch indexing system includes all of the tools an archive needs to set up and manage projects, organize groups of volunteers, and perform all of the tasks necessary to create digital browse aids and searchable indexes.

Group Administration

In the admin portal (accessible from any computer with Internet access), group administrators can create and manage groups of volunteers. Groups can be set up so that any user in the

system can see the group and request to join. Or, they can be set up so group administrators send invitations to only those users they choose to include. Groups can also choose to be associated with specific projects, or they can work on any open project in the system.

Group administrators can manage membership in the group, invite other users in the group to be administrators, send messages to group members, and see reports on the work done by members of the group.

Project Administration

Also in the admin portal, project administrators can create and manage projects. Projects can be set up as indexing projects, linking projects, or digital browse aid projects, depending on what tasks the project administrator wants volunteers to work on. The project setup process is very simple, yet there are many options the administrator can choose to make sure the appropriate guidance and restrictions are in place for those working on the projects.

Similar to the way groups are managed, project administrators can set up their projects so they can be seen by all and are open for any volunteer in the system to work on, or they can restrict access to only members of specified groups.

As the project goes through its lifecycle in the system, the project administrator can see progress reports and manage specific batches that volunteers send back for various reasons.

GedcomX Template Editor

All projects in the system are based on an open standard for storing and sharing genealogical records called GedcomX. For each record type in any project in the system, the project administrator creates (or modifies) a GedcomX template that defines what pieces of information (fields) are included in the record, relationships between those fields, what types of values are acceptable for each field, etc. The project administrator can also include project instructions, field helps, and sample images to help the volunteers be consistent in the way they complete their work.

Once a GedcomX template has been defined, it can also be exported for translation and then the translated versions can be imported back into the system so the project can be worked on by volunteers in different languages.

Ingest Tool

As part of the project setup process, the project administrator uses the ingest tool to ingest digital images and define batches of work for volunteers to work on. If the project is a linking project, the ingest tool also allows the administrator to import previously-indexed records and map the data to corresponding fields in the GedcomX templates associated with the project.

Indexing Tools

The FamilySearch indexing system includes several different tools that volunteers can use to index records. The primary tool indexes use is a browser-based indexing client. The volunteer simply signs into the website, selects a project to work on and begins indexing right in his or her browser. While the volunteer will typically have a constant connection to the Internet, the tool also has an offline feature so the volunteer can check out a batch and work on it even if there is no Internet connection.

FamilySearch has also developed a mobile indexing app for iOS (iPhone, iPad) and Android devices. This tool provides even more flexibility and allows volunteers to index small image snippets anytime, anywhere.

Linking Tool

Sometimes, an archive may have already created an index of a collection using some other method. In many cases, the specific records in those indexes were not linked to the corresponding images in the collection. One of the valuable elements of a search index is the ability to quickly jump to the source image once a record is located. The linking tool allows volunteers to link records from previously-indexed sources to the corresponding image in a collection. The volunteer looks at various images from a collection, identifies the records that belong on each image, and links them from the list of records that were previously created.

The original index may have been incomplete for one reason or another, so the linking tool also makes it possible for volunteers to add records and link them to an image during this same process.

Digital Browse Aids

While a searchable index of every record in a collection is the most efficient way for researchers to find records, it is also the

most expensive to create since volunteers must transcribe much of the information on each image in the collection. As a result, FamilySearch and some archives choose to simply publish the images in certain collections online and allow researchers to browse through those collections image by image, much like a microfilm reader, but from the convenience of any computer with an Internet browser.

If the project administrator chooses to publish a collection as a browse-only collection, it is still helpful to provide digital browse aids to help the researcher quickly jump to the desired part of the collection.

The digital browse aid tool simply allows volunteers to highlight multiple images in the collection and tag them as a subsection of the collection. Typically, these tags represent a period of time, location, etc., depending on the collection type.

Developer Access to the System

One of the key components of the new system is a formal API that third-party organizations can use to integrate with the new system. This API is provided as a module of the FamilySearch API. The indexing API modules provide access to:

- Projects: a collection or group of records, such as Ohio birth records.
- Batches: one or more images from a project that someone can index in 30-40 minutes.
- Images: a digital document page from a project.
- Records: the name, date, place, or other information copied for an individual from an image.
- Users: the volunteers who index or arbitrate records.
- Groups: societies or other groups who have users associated with their group and track statistics and reports. Some groups provide training and support for group members. Some groups have specific projects or goals they would like to achieve. A group administrator manages the group.
- Invitations or requests: to join a group or participate in indexing work.
- Reports: usually for a group or project.

If you are interested in building onto or integrating with the FamilySearch indexing system, you can join the FamilySearch Developer Network (devnet.familysearch.org) and become part of a growing community of family history developers. There you can find all of the API documentation with sample code snippets and gain access to the API reference system to test your code. With this developer access, you could:

- Design a mobile app that provides volunteers a fun or game-like way to help index records.
- Integrate project progress reports or group statistics on your website.
- Create your own browser-based indexing client.
- Provide more custom reports for your archive or group.
- Quickly set up your own projects and administer them, including custom skinning (branding it as your own) of administrative and indexing screens.

Deployment Options

The indexing system is built on a Cloud infrastructure, which means multiple instances of the system can be quickly deployed in

various data centers around the world. Each instance can also scale as necessary to handle the projects and volunteers it supports.

There are two supported deployment options available to meet the needs of FamilySearch and our partnering organizations, regardless of your size or location:

- **A FamilySearch-branded and hosted system** that allows both FamilySearch and affiliates to set up and administer their own projects with a single pool of volunteers that work on all projects. Volunteers can join multiple groups, and groups can focus on specific projects or any public projects. Projects can be set up as public or private (only visible to certain groups).
- **An affiliate-branded system hosted by FamilySearch.** The affiliate would set up and administer their own projects and recruit and manage their own workforce. FamilySearch would host and support the system.

The architecture of the system provides a lot of flexibility in terms of deployment options. If neither of these supported

deployment options work for your organization, we can work with you to meet your individual needs.

Conclusion

Searchable indexes and digital browse aids play an essential role in providing easy access to an archive's records. By using the tools FamilySearch is making available to archives, you will be able to serve your patrons more effectively and efficiently. We invite you to take advantage of these free tools and let us help you preserve and publish your records.

Author Biography

Scott Flinders is the Sr. Product Manager at FamilySearch responsible for the FamilySearch indexing product. As Product Manager, he is responsible for the overall direction and customer satisfaction of the software used by indexers throughout the world. Scott has more than 14 years of product management experience in the software industry, including companies such as Novell, Iomega, and NTT Communications.