

DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

Table of Contents

Problem Statement.....	1
Solutions.....	1
Ignorance is Bliss.....	1
Your Family Tree Software.....	2
Roll Your Own.....	2
Information Management Tools.....	2
Metadata.....	3
Example Software Tools.....	6
Photo Management Software.....	6
Using digiKam to Manage Photos.....	6
Document Management Software.....	13
Using Zim-Wiki to Manage Notes and Documents.....	14
An Aside About Planning and Note-taking.....	20
Notes about Note-taking.....	20
Planning - TODO Lists.....	20

Problem Statement

In the process of doing research we encounter and collect an assortment of many artifacts that either start out in digital form or that we convert to digital form. Any one of these artifacts most likely pertains directly to multiple individuals. The problem is how to most efficiently store and catalog the artifacts in a way that makes them readily available for use in multiple operating system environments.

In what follows, I will suggest some ways that this problem can be addressed.

By the way, this document is inspired by a presentation by D. Joshua Taylor in Pierre, SD, in the Spring of 2018. His presentation included a section on organization techniques, and this is what started me down the road to creation of this document.

Solutions

Ideally a researcher would like to have just one copy of a document and one set of information for an individual, who is uniquely identified. Since most documents and photos typically refer to multiple individuals, the trick is to store one just copy of the document/photo instead of having multiple copies strewn about. Following are methods that can be employed to deal with this issue.

Ignorance is Bliss

One (common?) way of not dealing with this problem is to make a copy of a multi-person artifact for each person/family represented in it so that it can be stored along with the other information associated with the



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

individual. For example, you might have a folder for each individual into which you place a copy of each artifact that is pertinent to the individual. This does not solve the problem and is clearly not efficient.

Your Family Tree Software

Good software provides for unique identification of individuals and artifacts, and it provides a mechanism for associating one with the other. This is not a bad way to solve the problem, but it can leave the artifacts in a relatively inconvenient state for analysis and retrieval. For example, how you might find all documents that pertain to two specific individuals is dependent on the capabilities of the family tree software. In other words, accessibility can be an issue.

Note that some software like Gramps addresses this problem by allowing assignment of tags to an object (image artifact, note, etc.). This provides useful options for filtering/querying these kinds of objects. However, I find that the user interface aspects of the feature are limited and so can be hard to learn and to use. That is why I mention the accessibility issue.

The fact that good family tree software assigns a unique ID to each individual can, however, be of value in coordinating documents that pertain to him/her. Storing metadata that includes the person IDs for people referenced in a photo or document makes it possible to store a single document and to retrieve it based on its association with one or more individuals. I have chosen not to use the person ID for this purpose, but in what follows I will note ways in which this could be done.

By the way, one way to avoid creating bogus person instances in family tree software is to only create an individual when you have a document that confirms existence. The family tree software then becomes the center of truth for the existence of people as well as their relationships to each other.

Roll Your Own

You can create your own database and interfaces. This an excellent approach that gives you full control, and, if the tools are chosen well, the results can be installed in multiple operating system environments. I think that most people will, however, not be comfortable with such an approach due to their lack of skill in using the required tools (e.g., spreadsheets, databases, user interface software, etc.).

Information Management Tools

Software tools exist that can be used to manage photos and documents. The advantage of this approach is that you don't have to roll your own custom solution. Also, if you choose an open source software application, you may also have access to a community of people with similar interests as well as a community of software developers to manage the software.

Such tools should provide convenient ways to assign and manage metadata that you associate with each artifact. The metadata should enhance your ability to find and isolate artifacts (documents and/or photos) based on ad hoc needs. For example, you might want to find photos of a particular ancestor in a location at a specific time.



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

Ideally the tools should be operating system agnostic, or at least there should be versions for Linux, Microsoft, and Apple operating system environments.

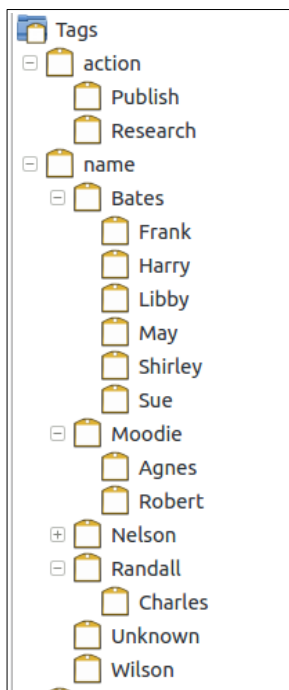
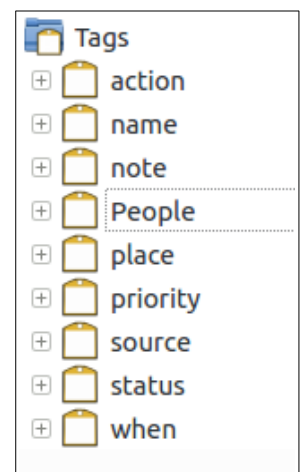
You will need to invest some time configuring the software and significant time assigning metadata values to artifacts. The rewards in terms of functionality and ease of use will, however, be apparent.

Metadata

Metadata is data about data. It can be recorded in many ways, but I have chosen to employ simple one-word tags. When possible, I suggest using hierarchical tags and consistency between tags on photos and other documents.

In a situation where hierarchical tags can be used, here is a list of possible top level tags from photo management software that I discuss later. These identify the type of information found at the next level down in the hierarchy.

Expanding this below shows what I have chosen at the next level in the hierarchy. (Note that this is just an example since your requirements made differ from mine.)



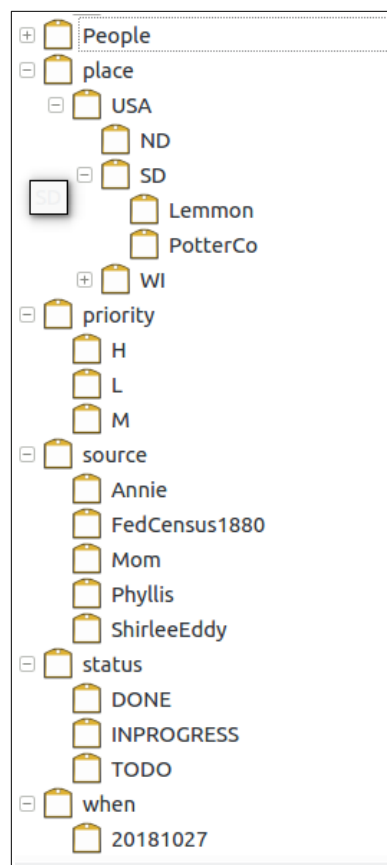
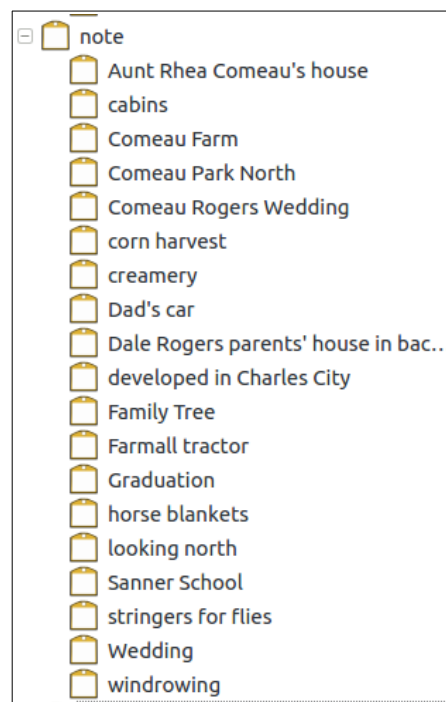
Under “actions” I have chosen two that pertain to what I am planning to do with the tagged object.

Under “name” there is another hierarchical list of surnames and their associated given names. Aliases can be handled by using multiple tags. Also, if you want to incorporate person IDs from your family tree software, it can be done here by appending the ID to the given name or by adding another level to the hierarchy below given name to hold the person ID.



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

I have chosen to use the “note” tag to indicate ad hoc information about the photo/document.



For now, ignore the “People” tag since that was added automatically by the software system that I was using to generate these pictures. I’ll explain it later.

The “place” tag is used to hierarchically record location information: country, state, county, city, etc. in this case.

I assign priorities simply as high, medium, and low using single letters.

“Source” is used to record the origin of the photo/document.

“Status” values are used to show where I am at with respect to the action noted above.

“When” is meant to help record when I acquired the photo/document or when it was originally created. This is situationally dependent.

DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

Even when hierarchical tags can not be explicitly employed as shown above, you can probably work around the problem using a scheme similar to the following one that I have chosen to deal with files in a software tool that only supports non-hierarchical keywords consisting of unbroken text strings. To solve the problem I format the keywords to represent hierarchical tags.

+ actionPublish	5
+ actionResearch	38
+ noteCivilWar	1
+ noteDNA	1
+ noteHollyHoward	1
+ noteLibrary	1
+ noteVirginiaHansen	1
+ placePACrawfordCo	1
+ placeSDCampbellCo	1
+ placeUKScotland	1
+ placeUSASC	1
+ placeUSASDEdmundsCo	5
+ placeUSASDEdmundsColpswich	3
+ placeUSASDFaulkCo	7
+ placeUSASDPotterCo	2
+ placeUSAWI	1
+ priorityH	5
+ priorityL	8
+ priorityM	29
+ seealsoDNA	1
+ seealsoMarshaSampsonGilliam	1
+ sourceAnnOHara	1
+ sourceDNA	1
+ sourceFedCensus1880	1
+ sourceMuseum	1
+ StatusDone	13

At left, you can see that I have chosen to use a text string in which a leading top level tag is shown in lowercase and the next level tag starts with an upper case letter (or a number). The general rule I use when assigning tags like these is to ask whether I can imagine a way to parse the string into components.

This scheme is not perfect, but so far it has been workable for me.

Note that there is one case where I have deviated from the previous high level tags. Not shown in the picture at the left is the fact that I use “surname” instead of “name” as a top level tag. This reflects a decision to not use full names or person IDs in these sorts of tags. (You’ll see why this works for me later.)

So, in the end, this scheme results in keywords that are composed of tags that are consistent, if not identical, no matter whether they are implemented in an explicit hierarchy or just in a single text string.



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

Example Software Tools

To further illustrate the use of metadata and available tools, here are two examples – one of photo management and the other of managing documents (other than photos) and notes.

Photo Management Software

DAM! - that is, Digital Asset Management - This is a term sometimes used to describe the processes for storing, manipulating, and in general caring for photographic data (i.e., pictures). A more detailed discussion of this topic can be found using the following url.

<https://docs.kde.org/trunk5/en/extragear-graphics/digikam/using-dam.html#keywords>

Also, I'll mention a simple, clear reference about photo organizing. -

Photo Organizing Made Easy, by Cathi Nelson, c.2017, isbn 978-1-946384-22-5 and 978-1-946384-23-2.

I found this book through a reference from the people behind the scanning software (VueScan) that I use (see <https://www.hamrick.com/>).

Using digiKam to Manage Photos

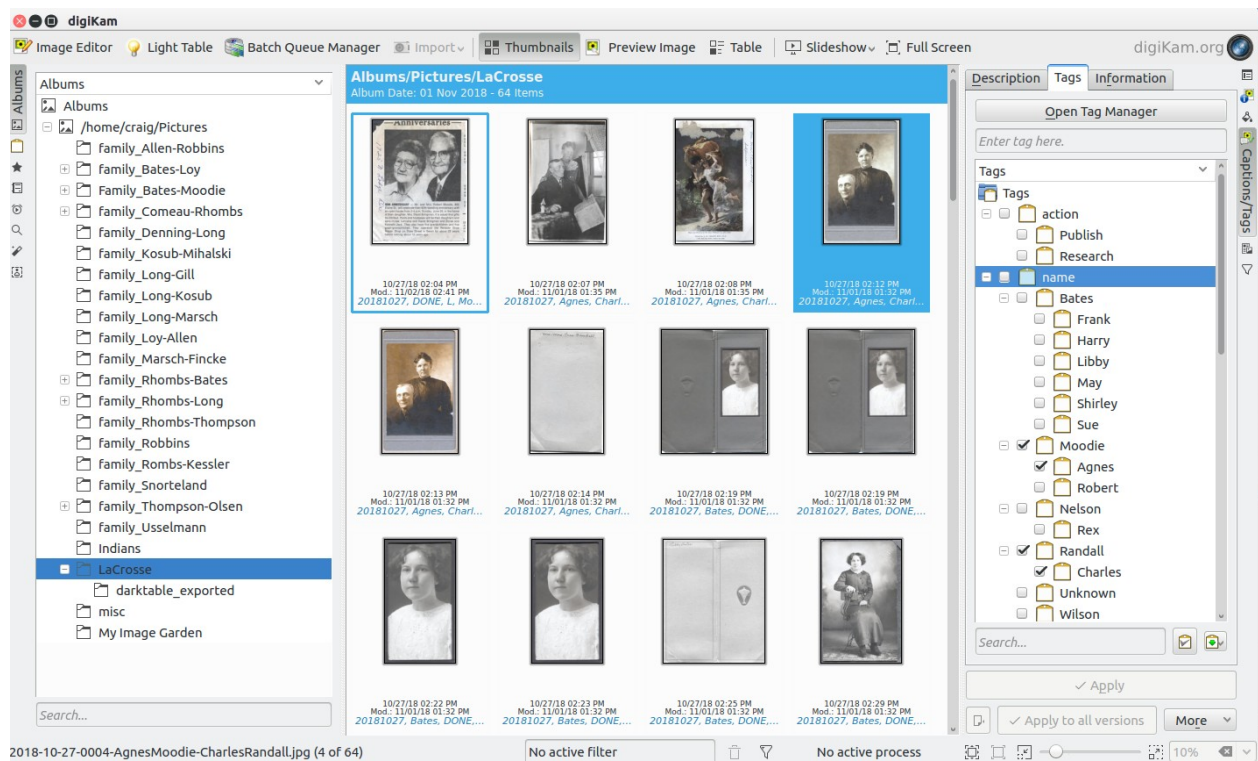
Because of its availability for multiple operating system environments and because of its usability, I have chosen a software product called digiKam (<https://www.digikam.org/>). The fact that it is an open source product also influenced me. Here is an article that explains some considerations in choosing similar software.

<https://photography.tutsplus.com/tutorials/how-to-choose-asset-management-software-for-your-digital-picture-archive--cms-24356>

All of my photos are stored in directories under my home directory named "Pictures". digiKam scanned this folder and discovered the photos there. Here's an example of how the main display appears afterwards.



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)



This is a three pane display. Notice that the list of folders is on the left, one photo is selected in the center (upper left part of the pane), and its tags are shown in the right pane. (Using digiKam I have added tags to some of the photos in a way that stores the tags in a local database as well as in each photo in an area of EXIF data. This is a standard associated with photos which specifies how extended information (like tags) is stored in the file that includes the actual photo. If a file is copied, the tags move with it.

Important features are as follows. The ability to to . . .

- select photos based on tag(s) or dates or using a timeline or keywords
- rate photos (used for planning work and culling)
- search for duplicate photos
- find people in photos
- for a given photo: display information about it, including tags, location, history, etc.; provide methods for editing exporting and printing.

There are a number of other features, but these are some key ones.

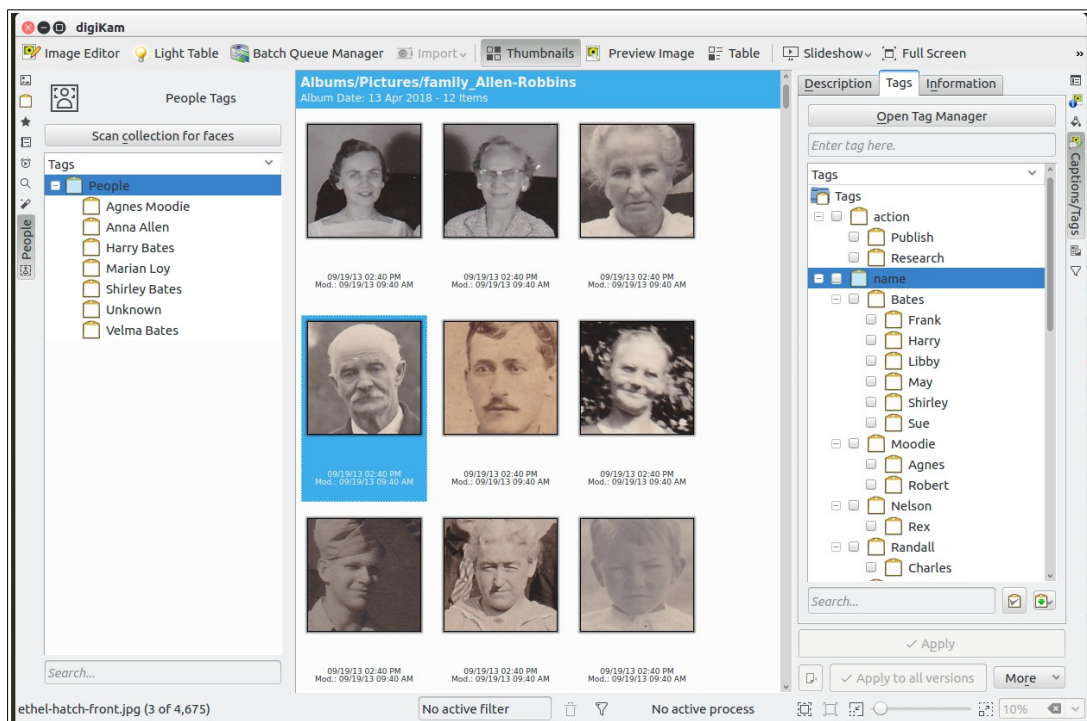


DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

For the purposes of this paper, the main feature of note is the ability to manage tags – apply tags and select photos based on tags.

In that regard, the ability to find people in photos is related. (The 4th bullet above.) When you use this feature, a new tag named “People” is automatically created so that you can add identifying information to each person found. Here is roughly how it works. (Refer to the picture that follows.)

- Start digiKam and use the menu on the left fringe of the display to select the “people” option
- Click on the “Scan collection for faces” button located above the “Tags” pane. This starts a process that results in a list of faces found in the center pane. This list is all in thumbnail form.
- Select one or more thumbnails in the center and then use the fringe menu on the right to select “Captions/Tags”. This shows what tags have been added to the selected photos.
- At this point you can change the tags on ALL selected photos using the right pane.
- On a given thumbnail you can also take the following actions: describe who this is (by entering a name or person ID), dismiss the thumbnail entirely from consideration. When you enter a person description, the text is recorded as a tag instance under the “People” tag, and it appears in the list of People tags in the list in the left pane.

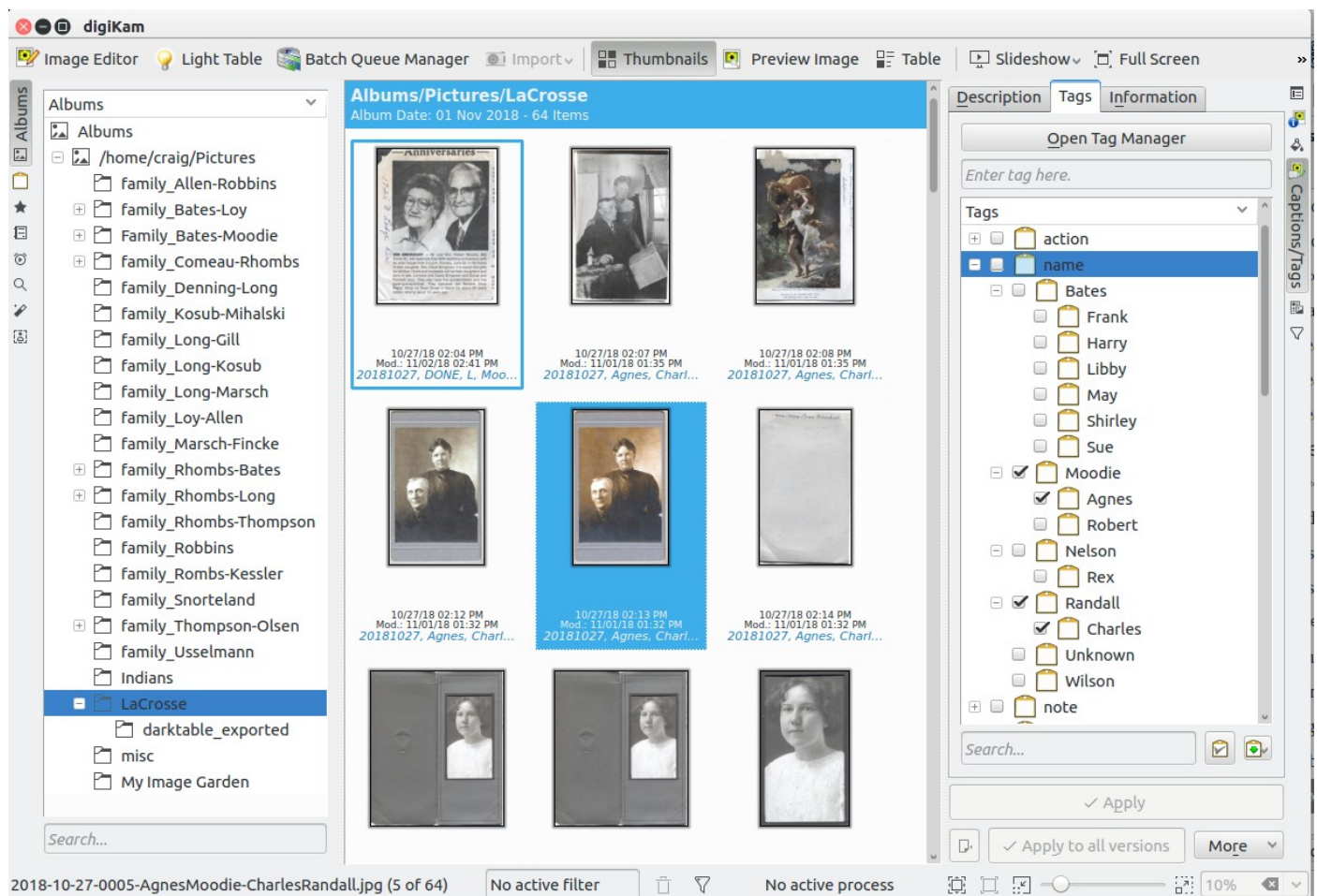


DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

This is a very intriguing feature and worthy of investigation. However, please note that some of the aspects of it are experimental. So, use it carefully! I would not rely on it to be the only way to tag photos, and that is why I have defined a separate “name” tag that requires manual assignment of values.

Here is an example of how I use the “name” tag to locate photos of people, independent of the facial recognition features discussed above. - This is built on my efforts at manually tagging each photo or groups of photos with name information. Once this is done, then it is a simple matter to query based on tags.

First, I select one or more photos, usually from a folder as shown below where just one thumbnail is highlighted (selected). In the right (“captions/tags”) pane I checked the boxes describing the two people in the photo (Agnes Moodie and Charles Randall).

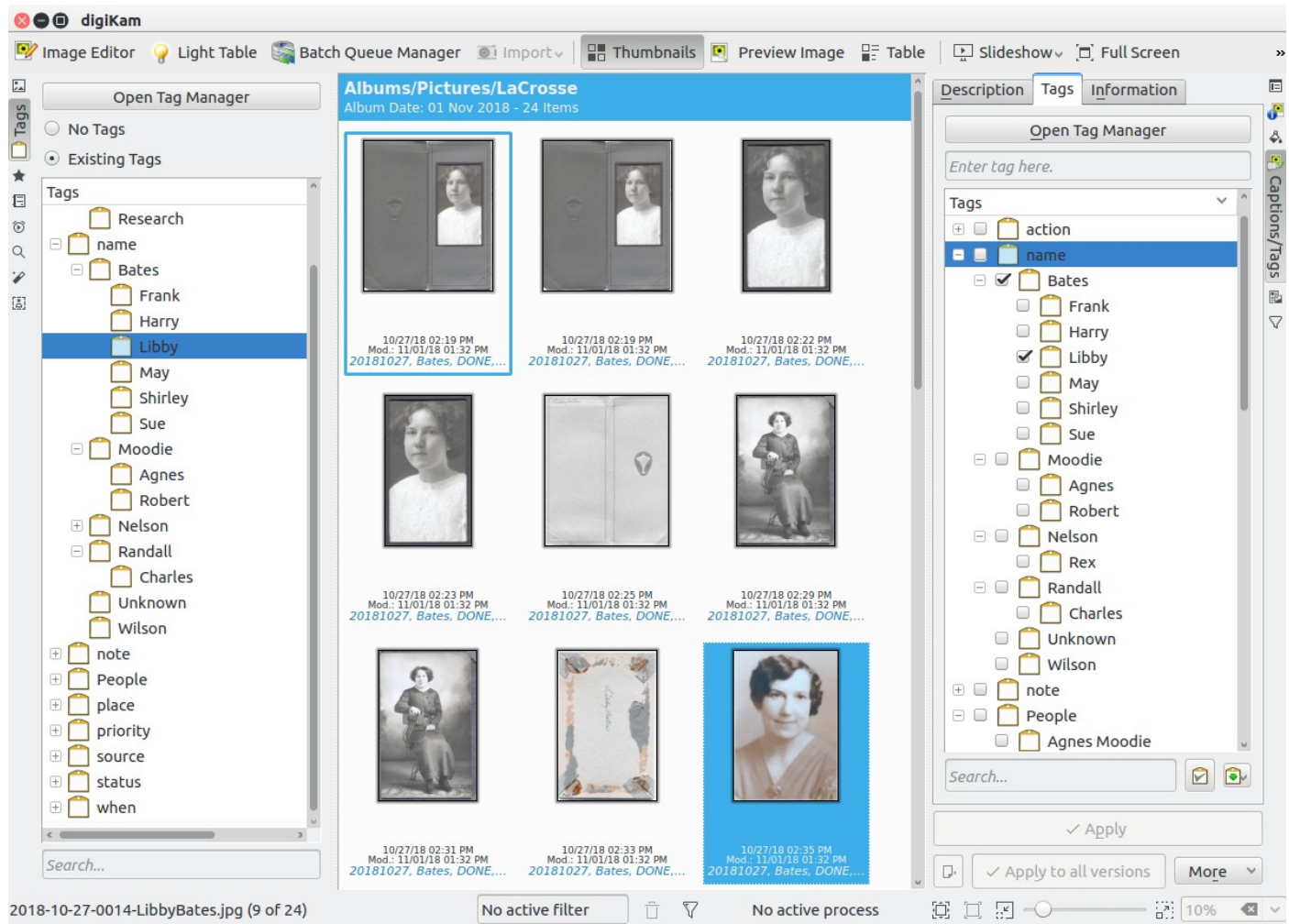


Once done, you can select photos based on the name tag(s) as illustrated below.



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

Here I have selected photos associated with Libby Bates.



I have selected the photo at the lower right. Its name, tags, and folder where it is stored are indicated in the center pane. Its tags are shown in the right pane.

It is important to know that once the EXIF data is defined in digiKam, it can write the values to the EXIF area in the actual photo. I have configured digiKam to do this, and to illustrate the result, following is information about the selected photo, as seen outside of digiKam.

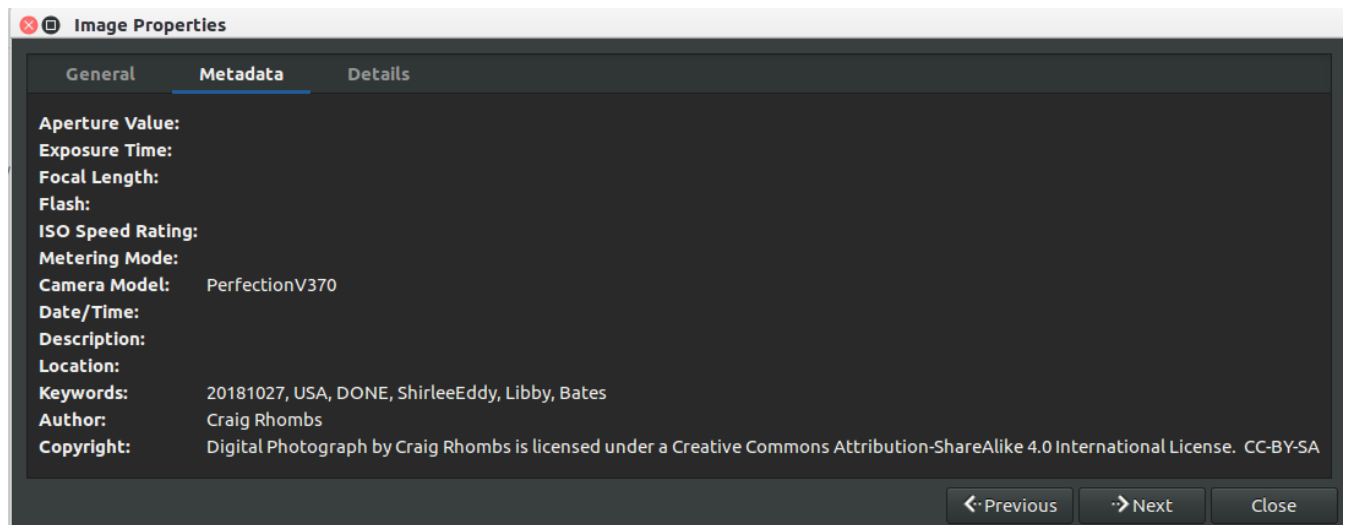


DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

Here is the photo as shown by another utility program.

Below is a listing of the metadata in the image file.
Note that it shows the tags created using digiKam as keywords.

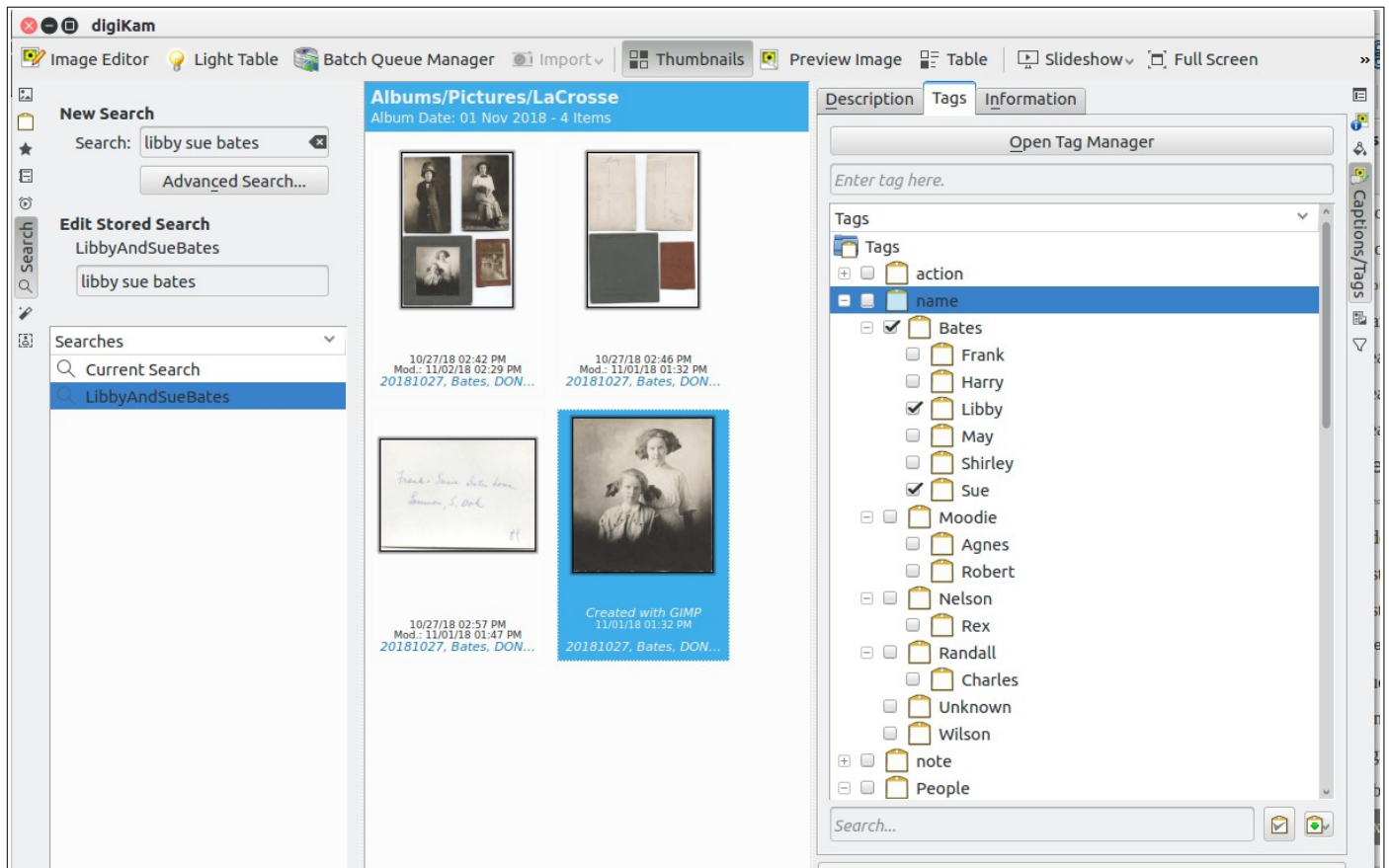
This metadata will follow the image file where ever it goes. So, you could export the image to another tool that sensed EXIF data and the original tags would be there.



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

The “Search” feature allows a user to do more complex searches from the left “fringe” menu. For example, you can search for photos based on multiple tag values. The example below shows the results of searching for Libby and Sue and Bates. (They were sisters.) I have saved the search and named it LibbyAndSueBates.

Tags of the highlighted picture are shown on the right.



Not bad!



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

Document Management Software

I have researched various ways to record metadata for files that are not photos. Some types of files allow for embedding metadata in a manner similar to that described above for photos. Unfortunately there are no universal standards that I can find. Consequently, I have had to turn to methods that store file metadata in structures outside the file itself. I have encountered two methods of note.

First, some file systems (like the “ext4” file system used in Linux systems) support extended user attributes. This method may be unique to the Linux world, but I will describe it briefly in the context of the Ubuntu-Mate distribution. - This distribution uses a file manager named “Caja”, which clearly supports the use of tags on files. Unfortunately, the default configuration of the ext4 file system in Mate does not support these extended user attributes and so the use of tags is in the end “blocked” by default. To make it work, the fstab file needs to be modified to define the (/home) partition where the files reside to support extended user attributes. I have not yet verified this, but the fact that this might be peculiar to Linux makes it less attractive. More research is required to confirm this. Here’s a place to start.

https://en.wikipedia.org/wiki/Extended_file_attributes

The second method is one in which a special software application stores metadata in its own database for file objects. I have chosen the Zim-Wiki product (<http://www.zim-wiki.org/>) to illustrate how this can work. Zim is another open source product that can be installed on multiple operating systems. Following is a brief description of how to use tags inside Zim-Wiki.



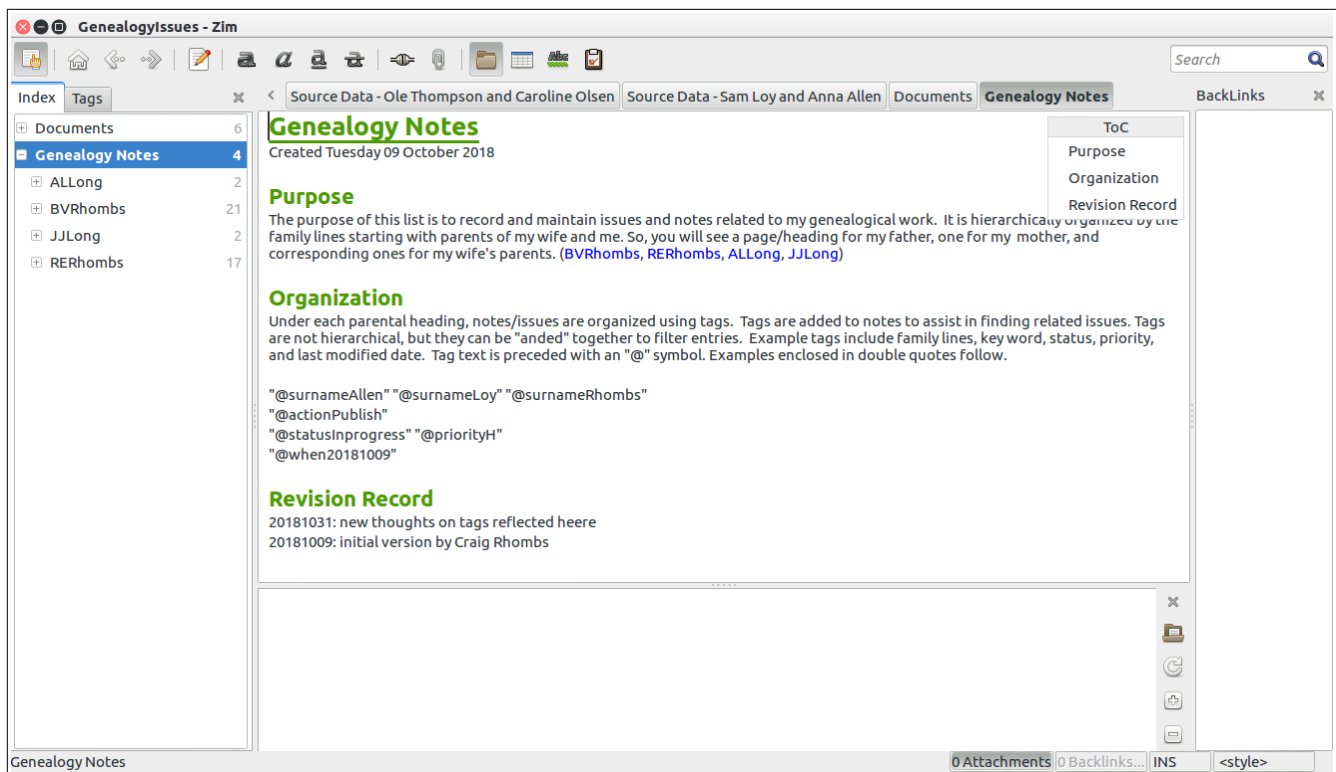
DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

Using Zim-Wiki to Manage Notes and Documents

I originally encountered the Zim-Wiki product in my search for simple open source software that I could use to manage my genealogy website (<https://281connections.us>). That use has turned out well, including interactions with the main developer.

When I started investigating ways to handle metadata in the form of tags for notes and documents, I remembered that the Zim-Wiki product was originally developed to help deal with note-taking. That is the short story about how I have arrived at this point.

Here is a picture of a display from the genealogy notebook that I created in Zim-Wiki.

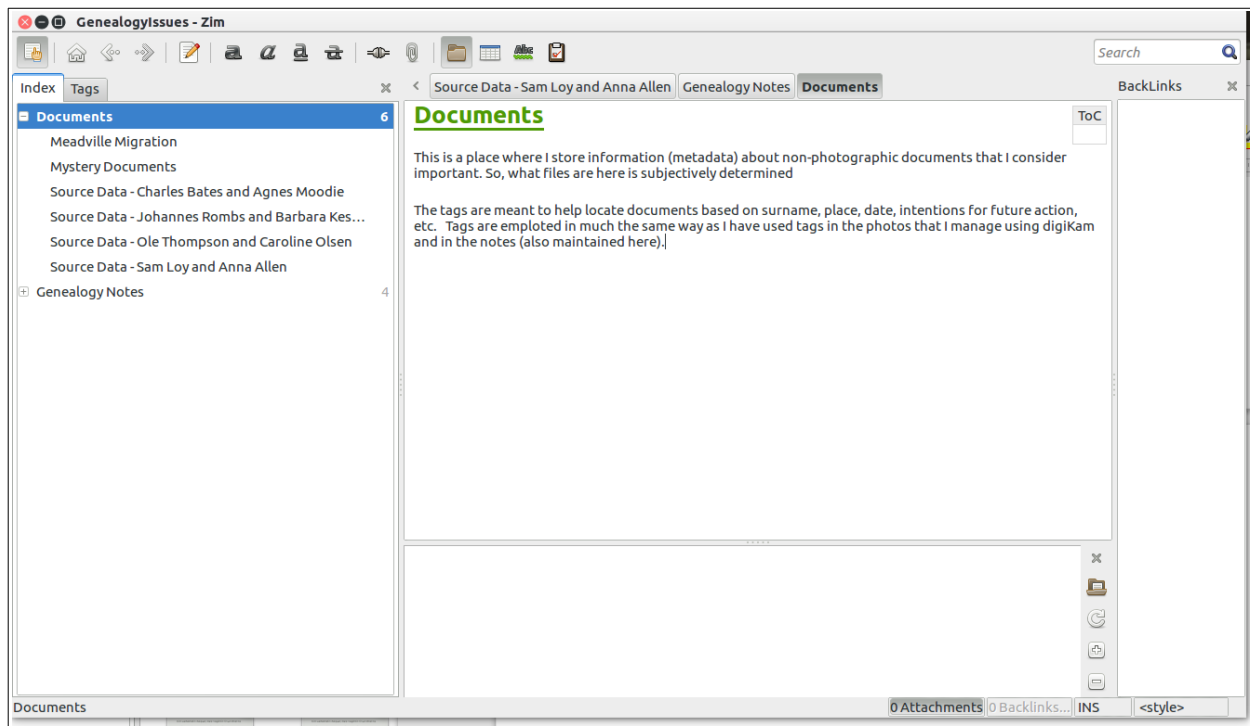


It includes a left pane that can be toggled between “Index” and “Tags” (guess where we are headed). A center pane shows the contents of the highlighted item in the left pane. I’ll ignore the right and bottom panes for now since they are optional.

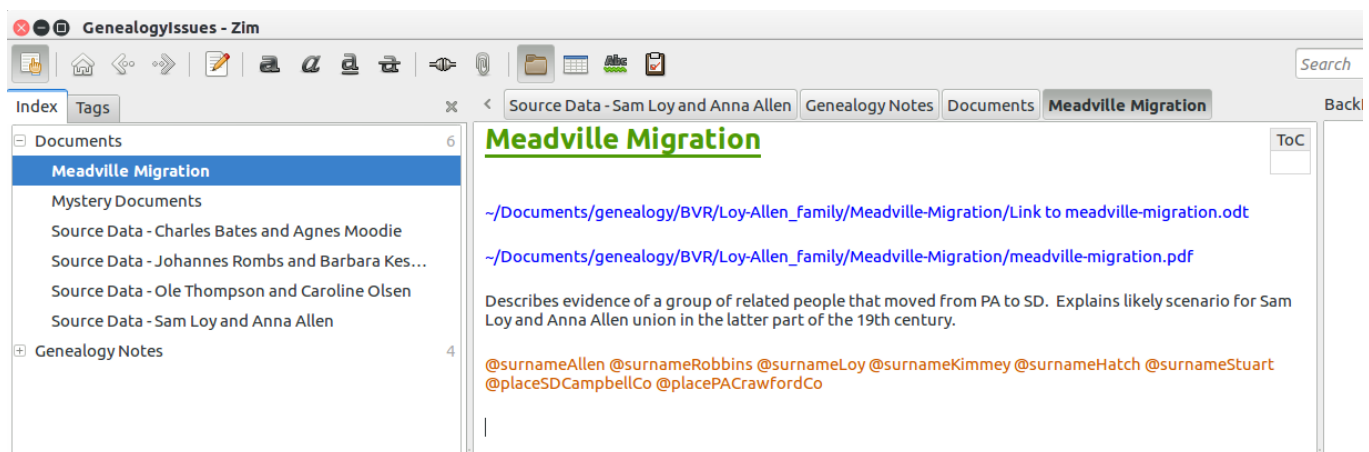


DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

You can see that I have created two main topics - “Documents” and “Genealogy Notes”. These share a common set of tags and a common index. The focus here is on metadata for documents, so here is a picture that shows what the display looks like when the “documents” item is selected in the left pane.



In the center pane there is a short description of the “Documents” item, which has six documents listed below it. Selecting the first listed document yields the following.



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

For this document (the “Meadville Migration”) there are two blue links to the actual document, then some text describing the document, and then some tags in yellow. The two links point to two forms of the document – an “odt” version and a “pdf” version created from the “odt” one. The actual documents files could be anywhere on my computer.

The eight keywords (annoyingly called tags in Zim-Wiki) are as follows. Each is preceded by the “@” character, and each is separated from the other with a space. Each is a complex text string with embedded hierarchical tags.

**@surnameAllen @surnameRobbins @surnameLoy @surnameKimmey @surnameHatch
@surnameStuart @placeSDCampbellCo @placePACrawfordCo**

You should recognize this form of the keywords/tags from the discussion above. - Since a keyword in this environment is a single continuous text string, a given tag is formed from leading lower case text that identifies the first part of the tag hierarchy (e.g., surname) and then text beginning with an upper case letter to describe the next component in the tag hierarchy (e.g., Robbins). This is the way that I have chosen to implement hierarchical tags in the Zim-Wiki environment.



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

Now let's see how the tags can be used. - Selecting the "Tags" tab in the left pane yields the following.

The upper part of the left pane shows a list of keywords/tags, ordered by the frequency of use in the notes and documents entries. This is a kind of tag cloud. (You may have seen similar name clouds in which the size of text reflects the frequency of occurrence.)

The lower part shows keywords/tags, ordered alphabetically.

The screenshot shows a Zim file manager window with the following content:

Tags Tab:

when20181009 actionResearch surnameRhombs
 priorityM StatusToDo surnameLoy surnameRombs
 StatusDone surnameKessler priorityL surnameAllen
 surnameBates surnameMoodie surnameSaenger
 placeUSASDFaulkCo StatusInProgress priorityH
 actionPublish placeUSASDEdmundsCo surnameLong
 surnameSnyder surnameRobbins
 placeUSASDEdmundsColpswich surnameThompson
 surnameOlsen surnameKosub placeUSASDPotterCo
 surnameKotzur noteCivilWar surnameWarner
 sourceAnnOHara sealsoDNA surnameHildebrand
 surnameLouth sourceFedCensus1880 noteDNA
 placeUSASC placeUSAWI placeUKScotland
 noteHollyHoward sealsoMarshaSampsonGilliam
 sourceDNA surnameRower surnameGill
 surnameWherry sourceMuseum noteVirginiaHansen
 noteLibrary surnameKimmey surnameHatch
 placeSDCampbellCo placePACrawfordCo
 surnameTupper surnameNicol surnameStuart

Tag List (Alphabetical):

+ actionPublish	5
+ actionResearch	38
+ noteCivilWar	1
+ noteDNA	1
+ noteHollyHoward	1
+ noteLibrary	1
+ noteVirginiaHansen	1
+ placePACrawfordCo	1
+ placeSDCampbellCo	1
+ placeUKScotland	1

Right Pane: Document titled 'Mead' with a description 'Describes Loy and A...' and a blue link '@surname@placeSD'.

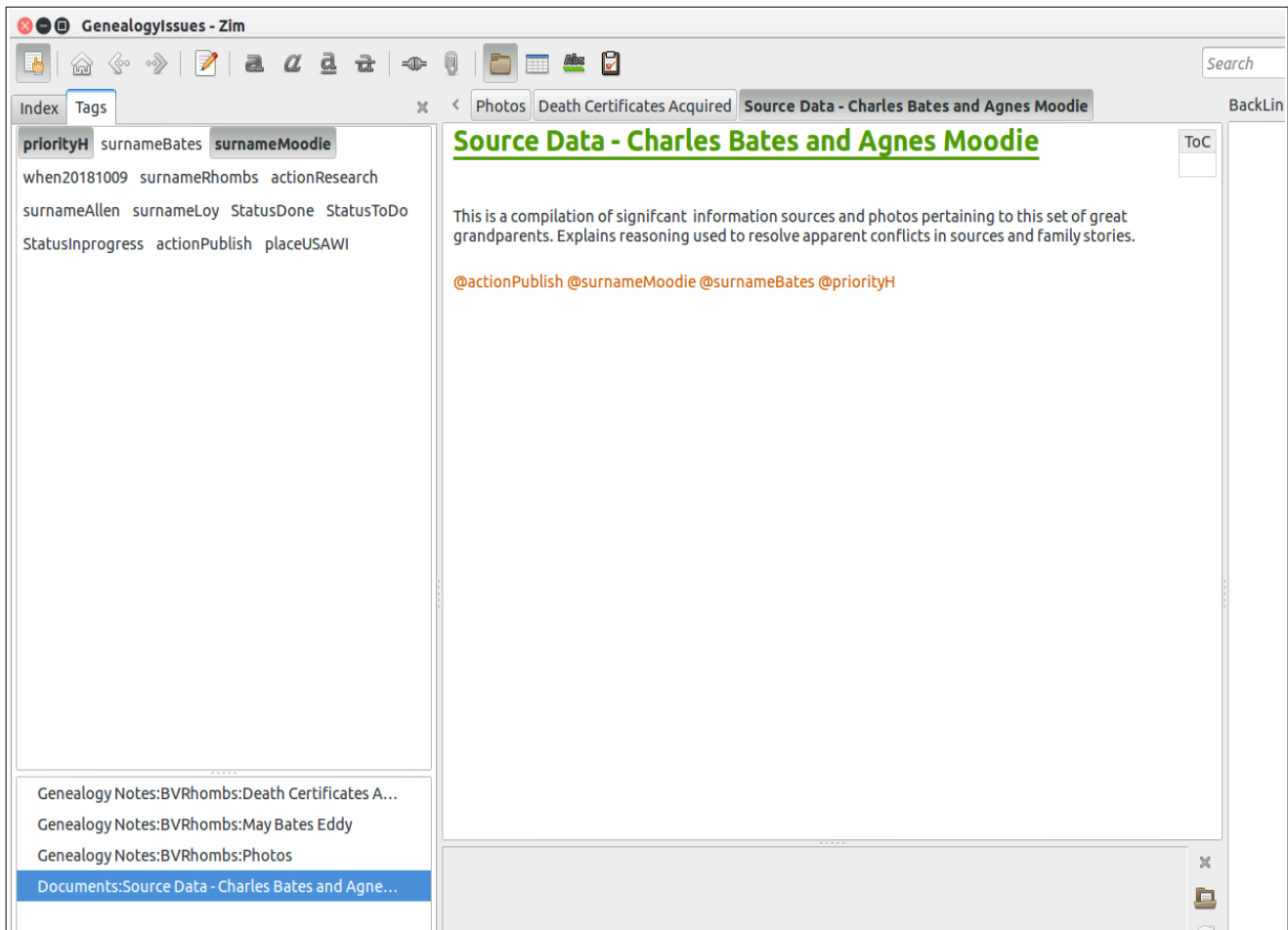
Bottom Bar: Documents:Meadville Migration



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

You have a couple of ways to use the left pane now. - You can select multiple keywords/tags from the tag cloud or you can expand one of the keywords/tags in the alpha list and select one.

For example, if I select high priority tags and those associated with the surname Moodie, I get a list of the matching notes and documents in the lower left pane. And, if I select one of the notes, I see a display like the following in which the actual note appears in the center pane.

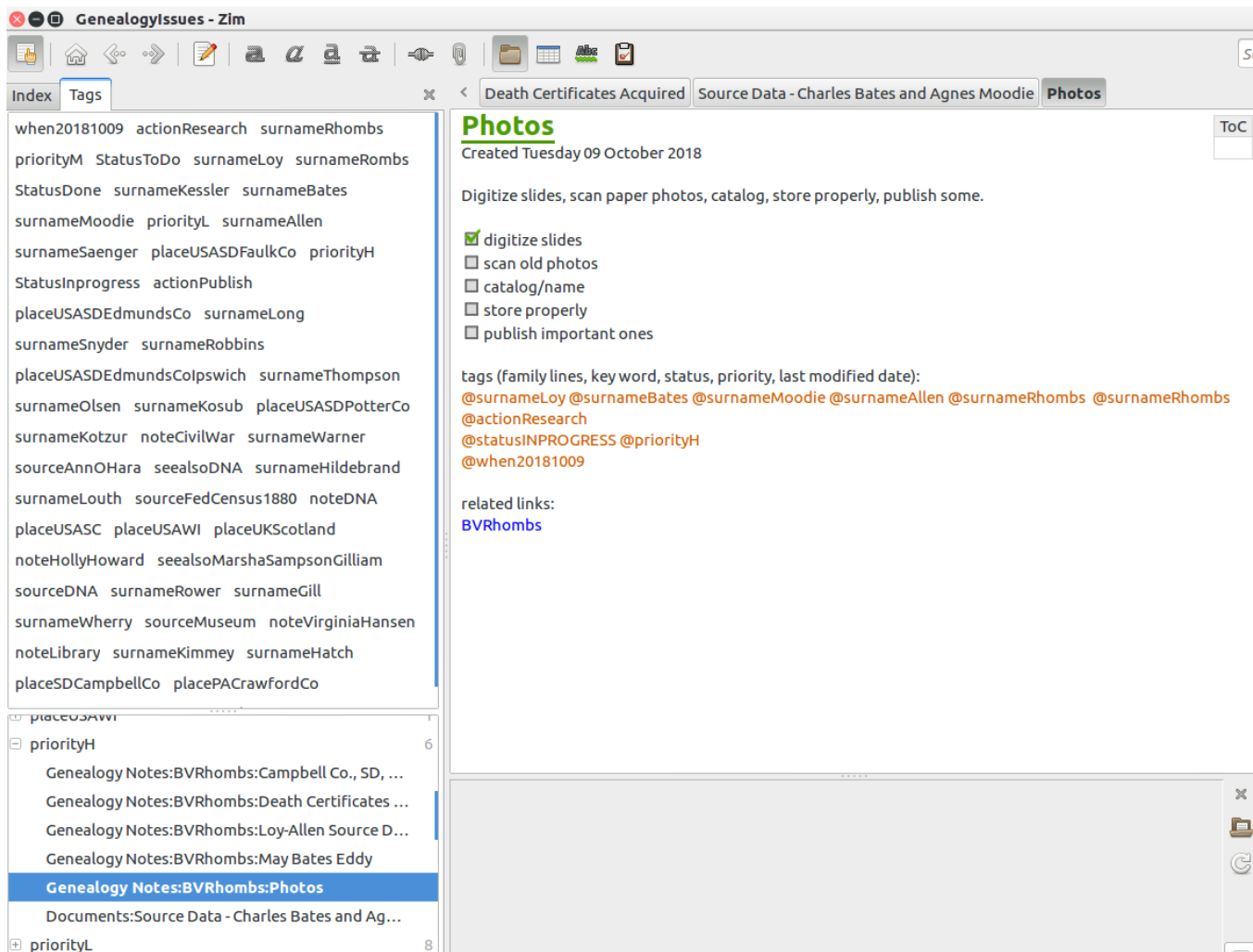


So, when selecting tags from the tag cloud, remember that you are filtering for those entries that match ALL of the selections simultaneously. (This is an “anding” filter process.)



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

In this second example, I have expanded the “priorityHigh” tag to reveal 6 items, consisting of one document and five notes. One note is selected that illustrates that you can have check-boxes in a note as well as a reference to another note or document entry.



So, I hope that you can see how the Zim-Wiki product or one like it can be used to associate metadata in the form of tags with notes and documents in a way that facilitates finding and maintaining them. Again, the advantages are (1) open source with associated support, (2) zero cost (although you really should donate to the team in some way if you make serious use of their work), (3) execution on multiple operating systems, (4) total control and ownership of data by you, (5) easy backup of your work, etc. Why pay more for less?



DRAFT - Thoughts About Organizing Genealogy Artifacts (documents, notes, photos/images in digital form)

An Aside About Planning and Note-taking

In the process of doing research, planning and note-taking are key activities. Here are some ideas that may help increase effectiveness.

Notes about Note-taking

There are (experimental) note taking features associated with the Firefox browser that may be helpful. I have used them to take rough notes that can later be transcribed to my more “official” lists mentioned above. I like this feature because there is a mobile aspect to this, so I can take notes on my cell phone in almost any environment and later can see them in a browser when I am at my desktop/laptop where I do the follow-up and transcription work. See <https://testpilot.firefox.com/experiments/notes> .

Simplenote is another application with similar characteristics that may appeal to you. See <https://simplenote.com/> .

Make sure that you understand what is secure and what is not. For example, the Firefox feature mentioned above encrypts the notes. I don’t believe that Simplenote does this. In general, it is a good idea to consider any on-line/web-based notes as public information unless you know for sure that this is not true.

College students have a significant need to take notes. Cornell University has described a note-taking method that they recommend, and it makes sense to me. You may find these references useful. The principles are pretty simple.

<http://lsc.cornell.edu/notes.html>

<https://youtu.be/WtW9IyE04OQ>

<https://youtu.be/pZgMpjjgCRA>

Planning - TODO Lists

Here’s a reference to a list making method that relies on two simple things – for each task, indicate a priority (like High, Medium, Low) and an estimate of the amount of time required to complete it. Then sort by priority and figure out how to use your time on the planet. --

<https://www.mnn.com/green-tech/research-innovations/stories/to-do-list-time-management>

