Personal Electronic Records Appraisal and Management Plan

By and For Allison King (and hopefully the whole King household)

Pat Galloway
INF 389G
April 30, 2007

I was both excited and nervous to find out that I had to appraise and come up with a management plan for my electronic records. I was excited because I have a large number of electronic files and I want to make sure that I know what I have and how to make it last. On the other hand, since I am an avid computer user, and have been for many years, I create digital records on a daily basis. I was nervous about having to look at this large body of material and figure out what to do with it. Since I use computers so frequently and for so much, all of the different roles in my life have many electronic records created for them.

I was 8 when I first started using computers. At first I mostly used them to play computer games. Then in high school I started to use e-mail, and became active on the Internet and a local Bulletin Board System (BBS) called A Place With No Name (APWNN). In college I began to use word processing programs and increased my Internet activity. During this time I also learned how to create websites in HTML and computer programs in C and JavaScript. At some point during this time I got my first digital camera and started playing with photo editing and image manipulation. Once I entered the work force, I learned how to make MS Office programs work together and how to use Adobe PageMaker and other programs both freely or commercially available and proprietary to the companies I was working for. Most recently, I've been learning digitization practices and how to integrate the components of Apple's iLife suite.

The process of appraising and developing a management plan for my electronic records consists of a number of steps, as was outlined by the Personal Records Appraisal Method and the Digital Records Inventory handouts. In accordance with these handouts, this paper will discuss my electronic records in terms of my roles, the scope of the collection, my file folder directories, the types of files I the collection, the programs used to create and/or view files, naming and metadata conventions, ownership of records, e-mail and other purely online records, missing files and analog counterparts, multiplicity of locations, and my management and preservation plan for these records. By taking a detailed look at my instinctive electronic records management style, I discovered that while I had a general knowledge of my files and file structure, I had a larger variety of records than I thought that I did; a reasonably well organized file folder hierarchy, but with some inconsistencies; a lack of consistency when it came to file

names and metadata; records that only exist online; missing files do to both crashes and bad practices; and a lack of backing up and preservation practices. Since starting this project I have already fixed some of the problems with the folder hierarchy and started to implement the management and preservation plan outlined later. In the future I will continue to use the preservation plan as well as be more aware of where files are place and the metadata that is included with them.

Roles

As with most people, I have a number of roles that fulfill in my life. Although the six roles given to us in the Appraisal Method rubric are Family Member, Friend, Citizen, Individual, Student, and Employee, I quickly realized that many of the interactions and transactions that electronic create records are the same for family members and friends. Therefore, for this collection, I will treat the Family and Friends roles as generally the same. The Family & Friends and the Citizen categories can both be split up in to Real Life and Purely Online groups. Furthermore, there is overlap between many of the Friends and Individual activities because many of my personal interests and activities can be done either alone or in a social (real or online) setting.

For Family and Friends that I have real-life interactions with, there are a number of activities that produce electronic records. Most of these involve communication, such as e-mailing, listserves, newsgroups, and instant messaging. Also, I spend a great deal of time communicating with my family and friends through blogs, both by posting in mine and commenting on theirs, and through social networking and invitation sites, like Facebook.com, MySpace.com, and Evite.com. I also buy my friends and family gifts via the Internet and take digital photographs of them, especially my cats. I have noticed that one of the things that my friends and I like to do when we get together is either show each other things or have each other listen to things that are either on the Internet or that we've downloaded from the Internet. Furthermore, I have created Web sites for my husband and for my cats, and I developed a Web site for my wedding[1], which was primarily to help my family and friends find out about the details of the event.

[1] http://wedding.dragonandrose.net

Over the 10+ years that I have been online, I have met and befriended a number of people whom I have never met in person. However, I would still consider some of these Purely Online friends to be just as close as some of my real-life friends. With these friends all of our interactions are online, through the same communication and social networking tools that I use to interact with people that I know in real-life.

I play role-playing games (RPGs) with my husband and with my friends. While the table-top and live-action RPGs are generally played with paper and pencil, they can also create records that range from e-mails wondering when the next meeting is, to blog posts about the gaming session, to writing out the back-history of characters. In terms of Massive Multi-player Online RPGs (MMORPGs), the only one I play with any regularity is Puzzle Pirates. While my husband and some of my real-life friends play the game, most of the time I am either playing by myself or with my crew, the Pirate Monkeys. I have never met any member of my crew outside of the game, but I have become quite friendly with some of them through our online interactions. Also, when I first started using the Internet, I spent a great deal of time in the Webchat Broadcasting System (WBS) Star Wars chat room role-playing Star Wars. While I made a lot of friends there, some of whom I still keep in touch with through the means listed above for Online Friends, and probably generated lots of records, the chat room and WBS as a whole had pretty much died out before the start date of this collection.

Unfortunately, the only roles I can think of that I do as a Citizen is paying my taxes and bills, and voting. I wish there was more that I did, but there really isn't. However, at a Citizen of the Internet, or Netizen, I manage online communities, participate in discussion forums, and even occasionally edit Wikipedia. I debated making this one of my roles as an Individual, but I felt that since it impacts more than just me, and it's not really something that I do with my friends but something that could make a difference to many different people on the Internet, that I should categorize it as part of being a citizen of the online world.

As an Individual I have a large number of things that I do. In just going through my day-to-day life, I keep a calendar, check my balance online, blog, listen to music, and watch movies and TV shows. I like to take digital photographs and then edit them and put them online. As a pagan, I keep a Book of Shadows, research pagan things,

and I have helped to write rituals, including my own wedding ceremony. Also, I occasionally write short stories, as well as character histories and background information for the RPGs that I am in. I also enjoy playing Puzzle Pirates, which as mentioned earlier, can be played alone instead of with a group of people.

As a student, I write papers and translations, go to and take notes in class, create and give presentations, and read and do research with both digital and analog material for class, papers, and presentations. I also create Web sites using HTML, CSS, and Dreamweaver, and mark things up using XML and its subsets. Furthermore, I edit wikis, and digitize a variety of items. Also, I send and receive e-mails, both personalized and over listservs, related to class projects and information.

I have had a number of jobs since I started college. The common thing that I did in all of them that created digital records was send and receive e-mails and get paid via direct deposit. Other than that, different jobs called for the creation of different records, such as recording the details of a customer call when I worked as a Customer Service Representative for a telecommunications company or updating client addresses when I worked for an insurance company. When I worked for a Web site design company, I created and edited Web sites and images. During my employment at a credit union, I created a large number of records, such as mail merges, database queries, Web site pages, and digital photos. At my current job at the iSchool's IT Lab, I create hard-copy, Web based, and video tutorials, as well as teach classes and give presentations.

Overview and Scope

For the purpose of this appraisal and management plan, a liberal definition of "document" will be used. I will consider any file that has been created, downloaded, uploaded, or viewed on my computers, that is not program or assists in running one, to be a document, as long as they are not files that are frequently deleted in the natural course of my interaction with the computer. Therefore, items such as chat logs, papers for written classes, and music files will be considered documents, but Web browser cashes will not, as they get deleted automatically every time I quit my Web browser. The reason behind this for this is as follows: if a record gives evidence into the rational of a private citizen, then the documents that help to record her life overall have the most

value. While it can be interesting to log ever detail of one's online existence, such as Gordon Bell's "life-logging," currently I feel that I do not have the physical space to store all of that sort of information, nor am I comfortable with people being able to see all of that. I sites that I visit most often, such as LiveJournal.com and Web sites for classes, are documented through the bookmarks on my computer. I believe that this is enough to give evidence as to what online locations are important to me.

Although I have been using computers since I was eight years old, the collection I will describe in this paper started when I was a first-year at Smith College and received my very first personal computer, a Macintosh Apple G3 running OS8 and then OS9 with a 4GB hard drive. I named this computer "Howard the Bold," and he (I personify my computers) served me faithfully for many years. The only exception to this is when he crashed on the night of September 14, 2000, but this will be discussed later. Several years ago poor Howard the Bold was no longer functioning at the top of his class, and so a new (second-hand) 350 G4 Tower computer running OSX.3 with a 30GB hard drive was purchased and subsequently dubbed Howard II. I moved all of my records, including papers, music, and e-mails, to Howard II at that time. In December of 2005, I made my decision to apply to graduate school and in honor of that decision, I took out a no-interest computer loan (I was working for a credit union at the time) and purchased a PowerBook G4, "Elektra," running OSX.4 with a 100 GB hard drive. While Howard the Bold is no longer with us, his hard drive is still somewhere in my apartment as my husband generally refuses to get rid of computer hardware. Howard II did make the move down to Austin with us and still runs just fine, if a little slow. When I started using Elektra as my main computer, I moved some of the files from Howard II over to her: such as my school documents, the files I was working on to apply to graduate school, and some of my pictures. There were a number of documents left on Howard II that were not brought over because I did not feel they were needed on my laptop, such as pictures of my cat who had died, or because I didn't want people to find them on my laptop, such as some bad pieces of erotica I tried to write while in college.

The scope of this collection involves records that fit the definition outlined earlier that are primarily located on Elektra, with a strong secondary consideration of those that are on Howard II. I also have documents that belong in this collection which are on my husband Jonathan's computer (a home-made machine running Windows XP), the computers I use or have used at work at the IT Lab, eclecTechs, and the UMassFive College FCU. I also store files on the DrakNet server (my Web hosting service), LiveJournal, Google, my iSchool server space, the UTexas Webspace, and Smith College's Sophia server.

Directories

The main directories on both Elektra and Howard II are Desktop, Library, Documents, Movies, Music, Pictures, Sites, and Games. However, Games and much of the Library do not contain records as defined earlier. Also, while the Desktop does contain records, it is often a holding place for documents that have recently been downloaded or uploaded to the computer until they are moved to an appropriate folder or deleted due to not being needed. The file structure of Elektra, the Documents folder on Elektra, and the UT Austin folder in the Allison School folder in the Documents folder on Elektra are outlined in Charts 1-3, located at the end of this paper. In the course of research for this appraisal and management plan, I changed some of the file structure on Howard II so that it was closer to that of Elektra, although there are still a few differences which will be discussed later.

As you can see by looking at Charts 1-3, with a couple of exceptions, there is not much depth to the hierarchy of the folders on Elektra. Most of the time, folders are created within the main folder and do not contain their own sub-folders. This is especially evident in the Pictures folder, where there are many folders listed. From looking at the names of these folders a pattern emerges. Picture documents can be split up into one of two categories: digital photographs that were taken for a certain event, such as 'Maine 06-19-06' and 'nyrenfaire04'; or image files that are either photographs of subjects that are photographed many times, such as 'me' and or 'boris&natasha', or images from the web that are used for some other purpose, such as 'LJ pics' which is where images that I use to make online avatars are stored. Although the diagrams do

note that some of these picture folders do have sub-folders, those folders, especially in the one-time-event folders, are generally created by JAlbum, a program that makes online photo albums, and therefore the contents of these sub-folders are the same images that are in the top level of the sub-folder or are files created by JAlbum. While it is important to note that these albums have been created, in this case, it is the photographs themselves that are important because,/ as long as these files exist, the albums to display them can be re-created at any time.

There is a difference between the Pictures folders on Elektra and Howard II. On Howard II there is a folder in the Pictures folder called 'peoples', which contains both random pictures of my family and friends as well as folders of pictures of specific events that involved my family and friends. I believe that the reason that Elektra does not have a similar folder is that, since I bought Elektra, either I haven't taken enough pictures of or been to many events with family and friends, or I haven't edited those pictures and so the raw files are still in the 'new pics' folder on the Desktop or only on Jonathan's computer. However, I will probably make a version of the 'peoples' folder on Elektra at some point in the near future, because it makes logical sense to me to have one and because I want the file structures on the two computers to mirror each other as much as possible.

Folders within the main folders have been developed in two different ways, which is consistent with traditional archival observances. Many folders were created because an event happened and a folder was needed to organize the information, such as the various wedding folders, 'resumes', and most of the folders within the Pictures folder. However, a difference was noted when it came to the folders that were created for academic purposes. In this case a folder for each class was made before the semester started in anticipation for files that would be placed in it. However, sub-folders under the various 'Papers' folders are only created if there were multiple projects or parts of projects that I have been working on, which have many files. In this case it becomes easier to find things if they are in separate folders. This ends up creating a very deep hierarchy within the 'Allison School' folder, as can been seen in Charts 2 and 3. It is interesting to note that the folders from my undergraduate studies at Smith College generally don't have sub-folders (and usually have between 1-6 files in them) where as

the folders in 'UT Austin' were often prepared with sub-folders for 'Articles', 'Class Notes', and 'Papers.' This can be partly explained because at Smith my studies were mostly in Classics and Theatre, neither of which tended to produce many formal papers, and now that I'm studying Information Science there are many more papers, presentations, etc that I must prepare for each class. This theory is supported by the fact that the top level folders for each semester at Smith are labeled, 'Papers [Semester Year]', whereas the ones from UT Austin are just labeled with the semester and year. There is always an exception to the grouping of the current semester, but all of the folders for those classes will be grouped into a '[Semester Year]' folder when the semester is over and they are no longer being accessed several times a week, at the minimum. However, the stronger explanation is that when I went to Smith, from 1998-2002, the Internet was still coming into its own in terms of what professors could do when it came to materials for classes. Instead of having Blackboard or eResearves, where professors could upload reading materials (as UT does), professors would have us buy paper course packets of readings for the semester and hand out paper copies of the syllabi. Therefore, the only electronic files that I would create for classes would be papers that I wrote or in the case of computer classes, the assignments for those classes.

All of the folders that can be seen in the charts are manually organized, with two exceptions. The first exception is the 'iTunes Music' folder inside of the 'iTunes' folder. I set the iTunes program so that it automatically organizes my music into folders first by artist and then by album. On Elektra I currently have 191 artist folders, and each artist folder has at least one album folder. According to iTunes I have almost 1900 music and podcast files, most of which are in .mp3 format. In order to make sure that these files are put into the correct folder, the metadata on them needs to be very exact; I will be discussing metadata later. The other exception are records that are automatically created or downloaded by programs and placed into folders in the Library, such as e-mails, chat logs, and browser bookmarks. E-mails are located in the Thunderbird folder in the Library, whereas instant message chat logs, blog posts, and browser bookmarks and cookies are located in their respective program folders (Adium, XJournal, and Firefox) in the Application Support sub-folder. I have no problem with these programs

placing these files in these places, but when I back-up my computer, I need to remember to backup these folders and their files as well.

Types of Files

On Elektra, records fitting the definition laid out above take up just over 32GB of hard drive space as of around April 12, 2007. Out of 44 different file formats listed in Chart 4, the most common types of documents are .jpg (2789) and .mp3 (1840) files. This surprised me as I thought I would have more music files than image files. However, I have a large number of music files on Howard II that were not transferred over to Elektra. The next largest sets of records are .doc (275), .html (156), .gif (138), and .tiff (123) files. I was a little surprised that I only had 68 .pdf files. I thought I would have more than that, because I have so many readings for classes that are in .pdf format; I must save fewer of them electronically than I thought that I did. Other formats that should be noted are the .css and .js files (because they are integral to the Web sites), the variety of movie and music formats (such as .avi, .imovie, .wav, and .m4p), and the .txt and .rtf files because they include both writings and Web pages. Out of all of the formats listed, the .itc files can be easily ignored as they are Album covers downloaded from the iTunes store and as long as the music files have album metadata attached to them, then they can be re-downloaded without any problem.

I was surprised to find such a variety of file formats in the Document folder; however, most of the variety is in the 'Allison School' folder and its sub-folders. For a long time that folder was part of the 'Allison's Writing' folder, but before I started graduate school I moved it to the top level of Documents. I believe that this was a good decision as I do much more as a student than just write papers, as evidenced from the fact that there are 20 different types of documents within that folder and its sub-folders. As I was doing this project, I was debating if I should move some of the records so that they go into the top-level folder that they fit into, such as moving the .mov files in the Perseus folder into the Movies folder. However, I decided that it would be better to leave them where they are because, while they are movies, they are also part of work that I have done for school and are linked to the Power Point presentation that I created for that project.

Programs

Even though there are over 44 file-types in this collection, many of them can be grouped together into the following categories: word processing/MS Office, images, music, movie, and code. This large variety of file types were created and can be viewed with a large variety of computer programs, many of which have gone through several versions or updates.

The word processing/MS Office files (those with .doc, .xls, .ppt, and .rtf) were primarily created with Microsoft Office. The earlier ones were created with Office '98 for the Mac, then 2001. When I switched over to OSX I used Office v.X until Fall 2006 when I upgraded to Office 2004. Additionally, some of these documents were created on the comparable PC version of Office and moved over to the Mac, at which point they were checked for changes to formatting, and fixed if needed. At this time I do not have any other program that can read the MS Office proprietary files. However, the .rtf files can be read by any text reader, such as TextEdit, which was also used to create the .txt records. The .pdfs can be opened with Preview, which comes standard with OSX, and with Adobe Reader.

Image files, such as .jpg, .tiff, and .gif, can be viewed with a wide variety of image viewing programs, such as Preview and Web browsers. When editing or creating images, I generally use Adobe Photoshop 7 or CS, depending on which computer I'm working on. Photoshop also produces the .psd files which should only be opened in Photoshop or ImageMaker due to the way the documents are structured. Most recently I have started using OmniGraffle, since Elektra came with a licensed copy, to create charts and other diagrams. While the original files are saved as .graffle, I always save a copy as an image or .pdf as well.

Music files can be played with a variety of music programs, but are generally played through iTunes. Currently I am using iTunes 7.1.1 but, since I started using it in 2001, I have been consistently updating the program with new versions as they come out. The .band files were created with Garage Band and the .cdp files were created with Sony Sound Forge when I digitized *Really Rosie* for my sister, but all of these files also have .mp3 counterparts which are played through iTunes.

Movie files are most often played through the open source program VLC, version 0.8.6a, manly because in order to play the .avi files through QuickTime, I would need to download a specific codec and I don't know which one it is. However, some of the movie files can be played with no problem by their "home" programs, such as QuickTime for .mov and Real Time for .rm. The exception is the movie of my wedding which I digitized using iMovie HD 5.0.2 from a video that a friend shot at my wedding. At some point I hope to edit it and burn a really nice copy onto DVD for my family and friends, but until then, it can only be viewed with iMovie.

There are five main types of coding records in this collection: HTML, CSS, JavaScript, XML, and C. All of them can be edited with a text program such as TextEdit. However, the first three are most often edited with Macromedia (now Adobe) Dreamweaver 8 or 9. These three can also be viewed with any Web browser, as can the XML documents.

As can be seen from the versioning history of programs above, I usually just use different/newer versions of the same software program as they come out, doing serial conversions as needed. As always, there are a few exceptions, mostly in terms of Internet software. My first browser was Netscape, then I moved to Mozilla, and most recently to Firefox, although I have stayed with version 1.5 of Firefox instead of upgrading to 2.0 because I like the way that 1.5 works better. Through all of these conversions, my bookmarks and other personal settings were transferred over without a problem. My e-mail program of choice followed a similar route from Netscape Mail to Mozilla Mail to Thunderbird. Again, all settings and e-mails transferred without a loss, as evident by the fact I still have e-mails that go back to the day after Howard the Bold crashed in 2001, even through that was three e-mail programs ago. When I started this collection, I was using separate proprietary instant messaging programs, such as AIM, ICQ, and Yahoo! Messenger. In the summer of 2004 I switched over to Adium, which allows me to log into all of those separate messaging systems through one program. I do not remember if the chats from the proprietary programs were being logged when I was using them, but I could not find any on Howard II.

Naming and Metadata

By looking at the file structures on Charts 1-3, the lack of a consistent naming convention can be seen. Although files and folders are generally given descriptive names, how those names are written changes without any pattern. Some of the folder names use spaces, some don't; some times full dates are written out sometimes only the month or the year is given. Examples include 'anna's wedding 05-20-06,' 'McGuffy_images', 'halloween05', and 'resumes'. Even within the class's folders in my 'UT Austin' folder, 'articles' was capitalized this semester, but not last semester.

When naming photos in a series, they are usually given descriptive terms of what the picture is about and, if there is more than one picture with the same subject, then a number is attached the file name. For example, in the folder called 'memphis', the files are as follows: awesomehat.jpg, bluescity.jpg, graceland01.jpg, graceland02.jpg,…, graceland10.jpg, and memphissunset.jpg. Often times when I name photos, I use an underscore to separate out words, because I like to place copies of my digital photos online and therefore, by using Internet naming conventions such as all lowercase and underscores, it makes the uploading easier. Another place where these Internet naming conventions are used is, predictably, in the Sites folder, because this is where the Web sites that I have made or are working on are located.

Other files, especially in the Documents folder, are generally give descriptive names, often but not consistently with capitalization and spacing. For example, this document is titled 'Records Management Plan.doc.' Very few documents have different versions saved; generally I just any changes over the old version. The exception to this is the papers that have been written for my Understanding and Serving Users class; since we have a group of four people who are all making changes to the papers, versioning helps the group to keep track of the latest copy so that new edits are not missed.

Unfortunately, the records in this collection do not have much metadata attached to them. Generally, the only metadata that they have is the file name, their place in the file folder hierarchy, date created and last modified, and anything that is automatically attached to them by the program that creates them. In the case of Word documents, the

metadata tends to be the name of the file, the author, and the company. However, the title listed when one goes to the properties of the file tends to be the first line of the file, which is often not what the file has been named. Also, the author is usually listed as either "Athene Numphe" or "Athene Parthenos," which are two names that I generally go by online, and the company is listed as "Goddesses R US," which is something I randomly came up with ages ago when Microsoft Office asked for a company. For the digital photographs, the original files have metadata about the camera that took them and when they were taken, which comes from the camera when the files are uploaded. However, when I edit these documents, I generally use Photoshop's "Save for Web" feature and when I do that, the camera metadata gets removed from the files. The original photographs, with their original file names from the camera, are stored on Jonathan's computer. Currently there is nothing linking the descriptively named edited photograph to its original.

In terms of music files, I generally am more aware of and keep better metadata than for other files. Because of the way that iTunes organizes my music (as mentioned above), I have to make sure that there is good metadata on these files or it would be impossible to find a specific music document outside of iTunes. In iTunes, all music files have the name of the song, the artist, and a genre attached to them. Many of them also have the album and track number tagged as well. In the course of doing this project, I discovered that iTunes has a way of taking metadata that has been entered into the program and converting it to ID3 tags, the metadata tags that most mp3 music players can read. Using this feature I made sure that all of my music files have ID3 tags on them and I will continue to use this feature to add the metadata to my music files from now on.

Ownership

In going through the files I discovered that I had quite a number of files, especially in my Pictures, Music, and Movies folders, that were not created (or owned) by me. Many of these files had been downloaded off of the Internet. Many of the movie files, such as the episodes of Battlestar Galactica, get routinely deleted after they are watched. I do this because movie files tend to be very large. The Movies folder takes up

17.87 GB of space, even though it only has 51 files in it, and 10GB of that is from the iMovie file of the video of my wedding. Files in Pictures and Music are generally kept. While I do not believe that this is currently a problem, as I am not making money off of them, the fact that I do not own all of them should be noted. With images that are not digital photographs, they are generally downloaded because I either wanted to use them to make an avatar for an online forum or chat or to be the background image on my computer. Many of music files were not encoded from a CD that I own or purchased from the iTunes music store; the RIAA would probably be upset if it found out that I have them. At this time I'm not going to do anything about it because I enjoy listening to the music that I have. I will note that according to the program syncOtunes, only about 15 of the almost 1900 files are protected with DRM. An interesting case of ownership is that of the *Really Rosie* files. I digitized a version of the Off-Broadway Production soundtrack because there is no CD of it in existence. While I believe that this is fair use, as I'm not selling copies of it, the ownership of these files could be a tricky question, mostly because the tapes and records of the recording still exist even though they can be hard to find.

Until recently, all of my original records that I have in my collection were created purely by me. However, since I have entered graduate school, I have been collaborating with group partners, some of the files in the 'UT Austin' folder, such as those in the 'Papers' folder in 'Users,' are coauthored with other people. While the names of all of the authors appear in the text of the records, it is not usually noted in the metadata.

As mentioned earlier, there are a few other records that are not listed on the charts, but are in the Library folder. Adium, XJournal, and Firefox all have folders in the Application Support sub-folder. Instant message chat logs from Adium take up 17.8 MB of space and are located in the 'Users/Default/Logs' folder. Each different chat program has its own sub-folder and then each contact has their own folder with their chats. Individual chats are stored as .AdiumHTMLLog files, which naturally open in Adium, but can also be opened and read semi-easily with a text reader. Bookmarks and cookies are in the Firefox/Profiles folder. I have a lot of bookmarks listed in the 'bookmarks.html'

document. However, I really only use the ones on the 'Personal Tool Bar' and I really should go through the rest of the list and clean it up sometime; I just keep forgetting.

I've been keeping a blog on LiveJournal.com since Summer 2001. Since then I have made 1,332 entries in my main journal. I have also made numerous posts to community blogs and a few alternate journals. According to my LiveJournal Info page,[2] I have posted 3,508 comments and received 2,896, as of April 18, 2007. I recently found out that XJournal, the program I use for writing my posts, downloads my "history", a log of all my posts, into a .plist file located in the XJournal folder under Applications support. While it is easier to view this file inside of XJournal, I can also open it up in OmniOutliner as well as any text program. The XML isn't quite as nice to read as the XML that I could download month by month from LiveJournal, but I am content with it for now. Unfortunately, I know of no way to save the posts that I've made to the many communities that I belong to. However, I have noticed that if it is a particularly important post, I will usually cross-post it to my own journal as well, and therefore I already have a copy of it in the .plist record. I also do not have a way to save the comments I have given or received, but in general I believe that LiveJournal will do a good job of saving them, and if it doesn't, I'm not quite as attached to them as I am to the posts themselves. I am supposed to get e-mails of comments that I receive, but this service has been sporadic at best, so while I do have about 500 of these comments from others, going back to mid-2003, saved in my e-mail, it is not a reliable method of collecting them.

E-mails and Purely Online Records

E-mails are not listed in Chart 4, but each e-mail folder in Thunderbird has files for the individual e-mail folders, that I've set up under each e-mail account, that can be opened and easily read in a text program; each folder of e-mail also has a .msf file, which is only readable through e-mail programs. I currently have 41 folders in 8 e-mail accounts that have been popped into Thunderbird and take up 99 MB of space, although 3 of these e-mail accounts no longer exist. I also have 8 folders in my iSchool account, which is set up as an IMAP account and is therefore under 'IMAP' and not

[2] http://athene.livejournal.com/profile?mode=full

'MAIL' in the Thunderbird folder. There are .msf files for each of the e-mail folders, but these cannot be opened straight from the file. I have chosen to use IMAP, and therefore leave the e-mails on the iSchool server, because I access my iSchool e-mail from a number of different computers and in a number of different ways and therefore it makes it easier to keep track of. However, I will probably change to POP settings after I graduate and before the account is closed so that all of the e-mails in this account will be downloaded to my computer and therefore saved before the account is deleted in accordance with the iSchool's current e-mail policy.

Folders in my e-mail accounts develop in similar ways to that of folders on my laptop. Most of them are generated when e-mails come into one of my Inboxes that need to be saved yet there is no appropriate folder for them that already exists. Most of the folders are based either on sets of similar people, such as 'family'; listserves or groups, such as 'The Insider', 'austin pagan', and 'ssffs'; or events, such as 'wedding stuff', 'bills to pay', and 'Classes'. Each e-mail account has different folders depending on what that account is used for. For example, the e-mail account where all of my bills are sent has a 'bills paid" folder whereas my primary e-mail account has 'friends' and a 'LJ' folder. I like to use filters for listserves so that new messages are automatically moved into the folder for that list, thus leaving fewer incoming messages in my Inboxes.

I found it interesting, going through my e-mail, that I seem to have deleted all of the school-related e-mail from when I was at Smith, but kept all the e-mail relating to the two clubs I belonged to and was an officer in. I'm not really sure why I did this. I probably felt that I had finished with my undergraduate education, but I still had an emotional connection to the clubs and the people in them. It should also be noted that while I have e-mails that I received going as far back as September 15, 2000, in the past I periodically deleted all sent mail over a year old (although I currently have sent mail going back to January 1, 2004). I am uncertain whether I want to continue to deleting send mail periodically or whether I should keep the record of all of my sent mail. For now I think I will save them, unless I start to run out of space.

In addition to the e-mail accounts that I access through Thunderbird, I have an e-mail account through Google's GMail service. This is an online Web based e-mail

account, which I use quite heavily. Currently I am using 533 MB of space for my account (18% of my allocated amount). I use GMail not only for e-mail but also for sending files (because it has a 10 MB limit on attachments) and instant messaging through GTalk. I don't know how many messages and chats my account has, but it is in the thousands. Because of the amount of space given and nature of GMail's storage, I tend to keep messages in it that I would probably delete if they were in an e-mail account was stored on my computer. Instead of folders, GMail works with a labeling system. I currently have 14 labels, which, like my folders, vary in function from friends and family (such as 'meg' and 'mer') to school projects (such as 'users group') to financial (such as 'paypal/ebay'). While I do have the option of popping and therefore downloading my GMails, due to the amount of them and the space that they take up, I don't think that it would be feasible. I also do not wish to do this because I like having an online e-mail account that I can access anywhere there is an Internet connection. At the present I trust Google to keep my e-mails as private as possible. If this changes, then I will make a decision of what to do, but for now, I'm very happy with their service.

Because I conduct many transactions online, there are a large number of Web sites that have electronic records about me or my family. Some of these are bill-paying sites where I go to pay my bills, handling such things as the cable bill, my credit cards, and my car loan. Others are accounts at financial institutions, such as credit unions, where I can login to check my balances, pay loans, transfer funds, etc. I have other accounts, at sites such as Pay Pal, Amazon.com, Apple, and DesignerShoes.com, that have my personal information because I buy things through them. Also, I have online accounts on LiveJournal, Blogger, Puzzle Pirates, a number of forums (such as ABC.com), etc., some of which have my personal information and some which don't. In terms of security and access of these accounts, I trust these companies to hold my information in accordance with their privacy policies and therefore am not overly worried about the fact that they have these details about me. While this may be a little naïve on my part, I have so many other documents to worry about that I am content to know that I have these accounts and to just keep my eye on the news for possible security breaches.

Missing Files and Analog Counterparts

On September 14, 2000, Howard the Bold crashed. Although I was able to recover some files, I ended up loosing a lot of records that I had created since I started college in Fall 1998. This is reflected in the fact that the oldest files in this collection are from December 1999 and the earliest e-mails are from the day after the crash. However, in the 'Papers Fall 99' there is only one paper in the 'Mythology' folder and I know that I wrote three for that class. Also, I cannot seem to legibly open the 'Wonders of the World' paper in 'Art' from 'Papers Spring 00' in any program that I have. The one paper that I really wish that I still had was a radio play that I wrote for my writing intensive course my first semester at Smith. While I believe that I still have paper copies of these papers in a box in my closet labeled "Allison's Smith Notes," I may have brought some of them, especially the ones from my first-year, to my mother's house in New Orleans -- in which case they were probably destroyed in aftermath of Hurricane Katrina.

It should be noted that, although I took four computer classes at Smith, I have very few files for any of those classes because most of the coding, whether in HTML or C, was done directly on the school's comp sci department servers and I never thought to download a copy of them to my computer. Unfortunately, the files were deleted after the semester was over to make room for the next semester's class. I do still have the files from CSC 102 in the 'old personal site' folder in the 'Sites' folder because I moved the files to my old personal Web site[3] when I was done with the class. I also have a little bit of one program that I wrote for CSC 111 that I was having trouble with. Furthermore, Web site that Cat Stolz and I created as our second project for CSC 105 was preserved by the professor because he liked it enough that he wanted to show it to future classes[4]; if he ever decides to remove it from the Web, it will be lost to me unless I e-mail him and ask him for the files, which I am seriously considering.

About half of my electronic records do not have an analog counterpart. This is especially true of records such as e-mails, instant message chats, and Web sites. I do

[3] http://sophia.smith.edu/~aking
[4] http://maven.smith.edu/~nhowe/teaching/csc105/hall/ak-ts-sp2002/index.html

have analog counterparts for papers that I wrote for classes which I turned in and had returned to me. As mentioned above, many of them from my time at Smith are probably in the box marked "Allison's Smith Notes" and those from my time at the iSchool are in a hanging folder box in my closet in folders for the class, along with my analog notes from the class and any articles that I printed out that I wanted to keep. Many of my digital photos and images do not have an analog counterpart, although I did print out some of the ones from my wedding. While I may not have analog, CD, or DVD copies of all of my music and movie files, they do exist in the world and therefore, if I were to loose the digital files, I could always acquire another copy.

Multiplicity of Locations

One of the biggest problems with this collection as a whole is that it is spread out over many places. As stated earlier, some flies that are on Elektra are not on Howard II and vice versa and all of the original files for the digital photographs are on Jonathan's computer. Sometimes I end up writing documents for class on the computers in the IT Lab and then I (usually) e-mail them to myself over GMail, but I don't always download the attachment from GMail onto Elektra. Duplicates of documents, especially photographs and Web pages, are on various servers on the Web, including DrakNet, LiveJournal, and Picasa (a Google image sharing service). As long as I have backed-up copy of the documents that are on the Web, I am not worried about the possibility of them being taken down. This has already happened to me once; I was using Comcast as an ISP and had to cancel my service with them because I was moving. I had a number of pictures on their servers and, when I canceled my account, they took down those sites. Because I had copies of the pictures on my computer, I was able to put them back up on the Web in other places and just adjust my links to reflect this. I am also not worried about the music files on my iPod, named Sharon. As long as I have a copy of them on my computers, I am not concerned about loosing the music on her. It definitely would not be fun to put it all back, but it would be doable.

Just like the pictures and music records, I need to make sure that I have all of the final versions of the documents that I wrote for class on my computer. Therefore, at the end of every semester I will check to see if the final versions of the documents are in

their appropriate folders. If not, I will download them from GMail to Elektra and check the "last modified date" against the version on Elektra, keeping whichever one is the final version. I will also bring my laptop into work with me and compare the documents on Elektra and the iMac that I use when I'm at work. If the final versions are not on Elektra, then I will move them using either FTP or a flash drive. For the time being these documents will not be deleted from the iMac at the IT Lab, the iSchool server, or GMail as I have plenty of space in all of these locations. However, if space becomes an issue then I know that I can delete them from one source without worrying about permanently loosing the documents themselves.

For the most part, the records that I created as an employee were kept with the company when I left the jobs. Although it would show some of the work that I have done over my career, I'm not overly worried that they cannot be included in this collection. However, there are two exceptions. When I worked as the Marketing Assistant for a credit union I created many documents to help with the everyday running of the credit union and an instruction manual for how to manage these documents and the processes that went with them. While I don't really have a need for most of them, as they are very specific to the credit union, I did come up with a trick for how to make sure the leading zero in a zip code comes through during a mail merge. While I could easily recreate this document, it would be nice if I had the original. When I worked for a web design company, I created and maintained a number of web sites. Some of these sites no longer exist. While I can use the Internet Archive to give examples of these sites, it doesn't always have a good copy of them and I wish that I had the original files in order to be able to show my work in the online portfolio that I am currently developing. At my current job in the IT Lab, I make sure to save all of the important material that I've created for tutorials or teaching purposes to the Lab Staff Archive, as we are directed to do in our Records Management plan. I will probably keep a copy of the final projects on my own computers so I will have my own copies for my portfolio.

Management and Preservation Plan

In order to make sure that documents are saved, all records that fit the definition as laid out in this document will go through a number of steps. Documents that are currently in use (such as this one), will be periodically backed up on a flash drive and on the 30 GB drive of Howard II, so that work done will not have to be re-created should something catastrophic happen to Elektra before it is due. Extremely important documents may also be uploaded to the iSchool server for further temporary safety. At the end of every semester, all new documents, along with e-mails and the blog record will be copied into the appropriate partition of the 80 GB hard drive of Howard II, keeping as close as possible to the original folder hierarchy. Then, every year at New Years, all records will be copied to a 120 GB external hard drive, which is only to be used for this purpose. At the same time, comparisons to files already on the 120 hard drive will be conducted and random records will be spot checked to make sure that no data are being lost.

I have already started to implement some of these changes while working on this project. While I was looking at the file structure and records on Howard II, I copied some files over to him from Elektra. However, it is a slow process since some of the documents are already on both computers. As soon as the semester is over, I hope to spend some time moving the records so that everything is in concordance with the plan outlined above. My biggest concern is the music files, as I have edited or corrected a large amount of metadata of the music files that are on Elektra and I'm worried that the computers will not consider them the same file. This is especially true if I use a program such as syncOtunes, which is supposed to compare two iTunes Libraries and transfer files that are missing from each one. The files I care the least about are the ones in my Movies folder, as most of them I watch once and then delete. The obvious exception to this is the movie of my wedding, which is already backed up on the 80 GB drive on Howard II and on Jonathan's computer.

At the present, I do not have a disaster plan. I want to look into some of the online storage services that exist, but I haven't had the time as of yet. My current plan is put my laptop and the 120 GB hard drive in my backpack, grab my cats, and run.

As of now I am happy to continue to use serial conversion for my files or to use migration-on-demand, preserving them in the formats that they are currently in and opening them as needed. All of the major programs that I use I either have licenses for, such as Microsoft Office, or are open source, such as VLC. However should they need to be migrated to another format, different file formats have different requirements as to what is needed for preservation. In terms of papers and presentations created for class, formatting is as important as the content, especially in situations where there are tables, images, and other special formatting that help to give the paper its definition. In order to preserve this, I was considering converting all of my papers from school and a few other important ones, such as my wedding ceremony, to rich-text format and all of my papers and presentations from school and work to .pdf (since it is no longer proprietary to Adobe). At this time I don't believe that it is necessary to do this, as I plan to continue using Microsoft Office in the foreseeable future. However, I am aware that these options are available to me should the need arise. As for the files that can open with TextEdit, most of them are in an open source, such as XML or text and should last for a long time.

With images it is important to preserve the image as a whole. I don't believe that .jpgs will go out of style any time soon and therefore as long as they can be viewed with programs that come with operating systems or with Internet browsers, I'm not worried about most of my image files. I feel pretty much the same way about my music and movie files. As long as there is a way to play them, then they should stay in their current format. The only ones that I'm slightly more worried about are the two sets of files that I digitized myself: my wedding movie and the *Really Rosie* soundtrack. Although I could always digitize them again if I have the analog tapes, I probably won't always have access to the equipment. These files need to be checked every year during the large back-up to make sure that they are still accessible.

Over the course of doing this project I ran into a number of problems that I had to solve. Some of them, such as what to do with the movie files in the Documents folder, have already been addressed earlier in this paper. One of the other problems that I noticed is that here are a number of files that I have a number of copies of -- not just on

different computers, but also in different folders on the same computer. Most of these are pictures which are in the Pictures folder but have also been included in Webs sites and therefore are in the Sites folder. In this case, I think that it is acceptable to leave them in both places, because they belong to both places. Since the records are digital and I have the space, I am able to do this without a problem. If hard drive space starts to become an issue, then I'll look into alternative solutions. The other major problem that I noticed is that there are a number of sub-folders that have the same name. Records centered around my wedding are the most obvious case of this, as there is a 'wedding' folder in Documents, Sites, and Pictures. I was debating if I should group all of them together or leave them as is. I decided that the documents should stay separate from the pictures from the Web site, as they are now. Each folder represents different parts and activities involved in planning my wedding; although they are not in the same top-level folder, they are connected by their name.

Overall, this project has taught me that I have a lot of files in a lot of places and that I need to get a better control over them if I want to preserve them in the long run. Also, while I had a pretty good idea before starting this project of what and where most of my digital files on my laptop were, I made a lot of assumptions about what I would find in each of my major folders. Often times there were records of other types that were in the folders that fit in so perfectly that although originally I thought that they would not belong, I eventually decided not move them. I did do some rearranging as I found files that were part of a group, such as creating the South Park folder for my South Park movies, but most of that was fairly minimal. This project also brought into sharp relief the fact that my backing up processes were severely minimal to non-existent. I'm not sure when I got so lax about that. When I was at Smith I had a copy of all of my works-in-progress on the server and on a zip disk, especially after I'd lost so much when Howard the Bold crashed. I am glad that I was able to start backing up some of my files up and plan to do more of it in keeping with my plan very shortly. This project has been extremely useful to me, even though it took such a long time as I have so many records that fit my definition as outlined at the beginning of this paper. I feel that I have a greater appreciation for what I have and how to make sure that it lasts a long long time.

Chart 1

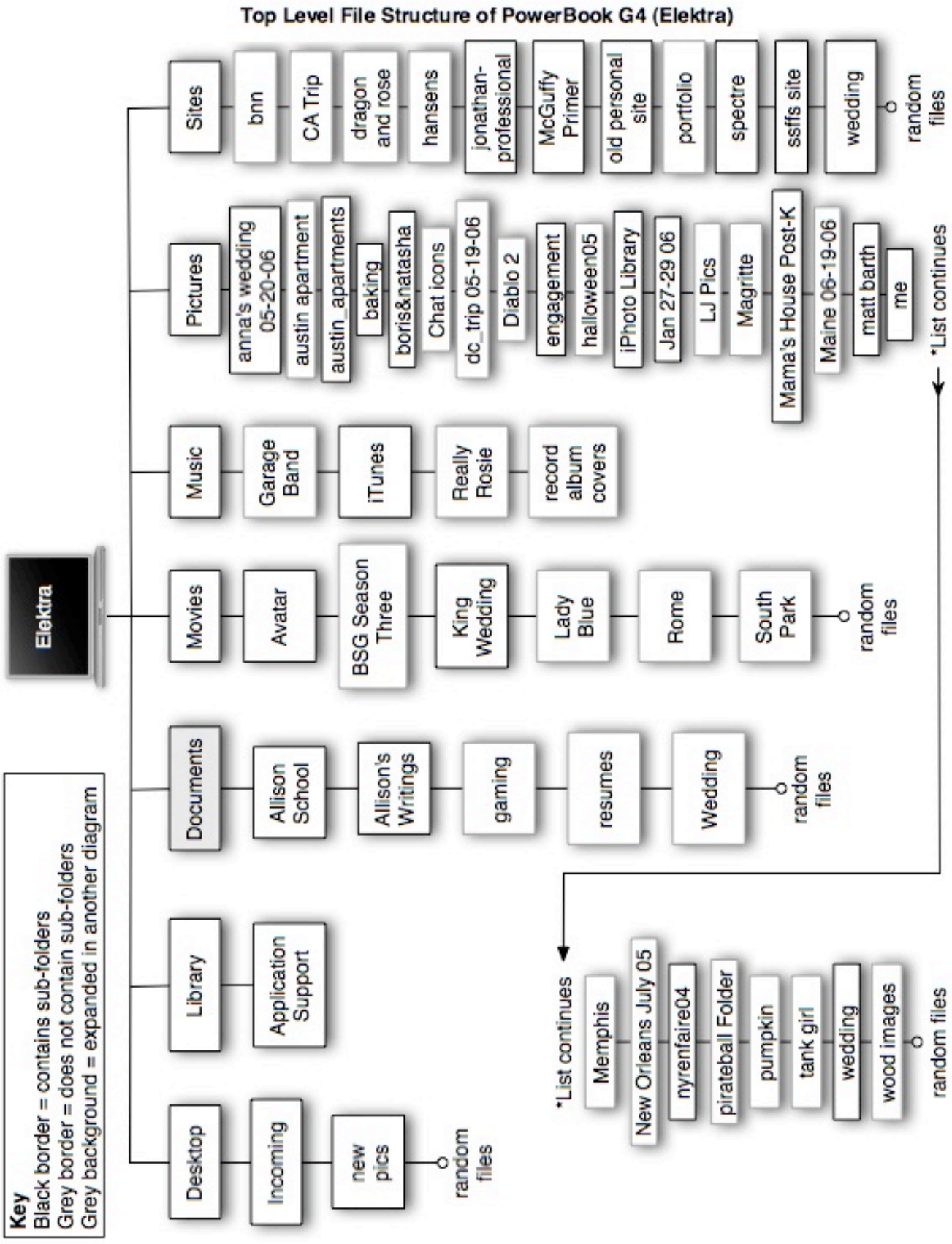
Top Level File Structure of PowerBook G4 (Elektra)
Key
Black border = contains sub-folders
Grey border = does not contain sub-folders
Grey background = expanded in another diagram
Elektra
Desktop
Incoming
new pics
random files
Library
Application Support
*List continues
Memphis
New Orleans July 05
nyrenfaire04
pirateball Folder
pumpkin
tank girl
wedding
wood images
random files
Documents
Allison School
Allison's Writings
gaming
resumes
Wedding
random files
Movies
Avatar
BSG Season Three
King Wedding
Lady Blue
Rome
South Park
random files
Music
Garage Band
iTunes
Really Rosie
record album covers
Pictures
anna's wedding 05-20-06
austin apartment
austin_apartments
baking
boris&natasha
Chat icons
dc_trip 05-19-06
Diablo 2
engagement
halloween05
iPhoto Library
Jan 27-29 06
LJ Pics
Magritte
Mama's House Post-K
Maine 06-19-06
matt barth
me
*List continues
Sites
bnn
CA Trip
dragon and rose
hansens
jonathan-professional
McGuffy Primer
old personal site
portfolio
spectre
ssffs site
wedding
random files

Chart 2

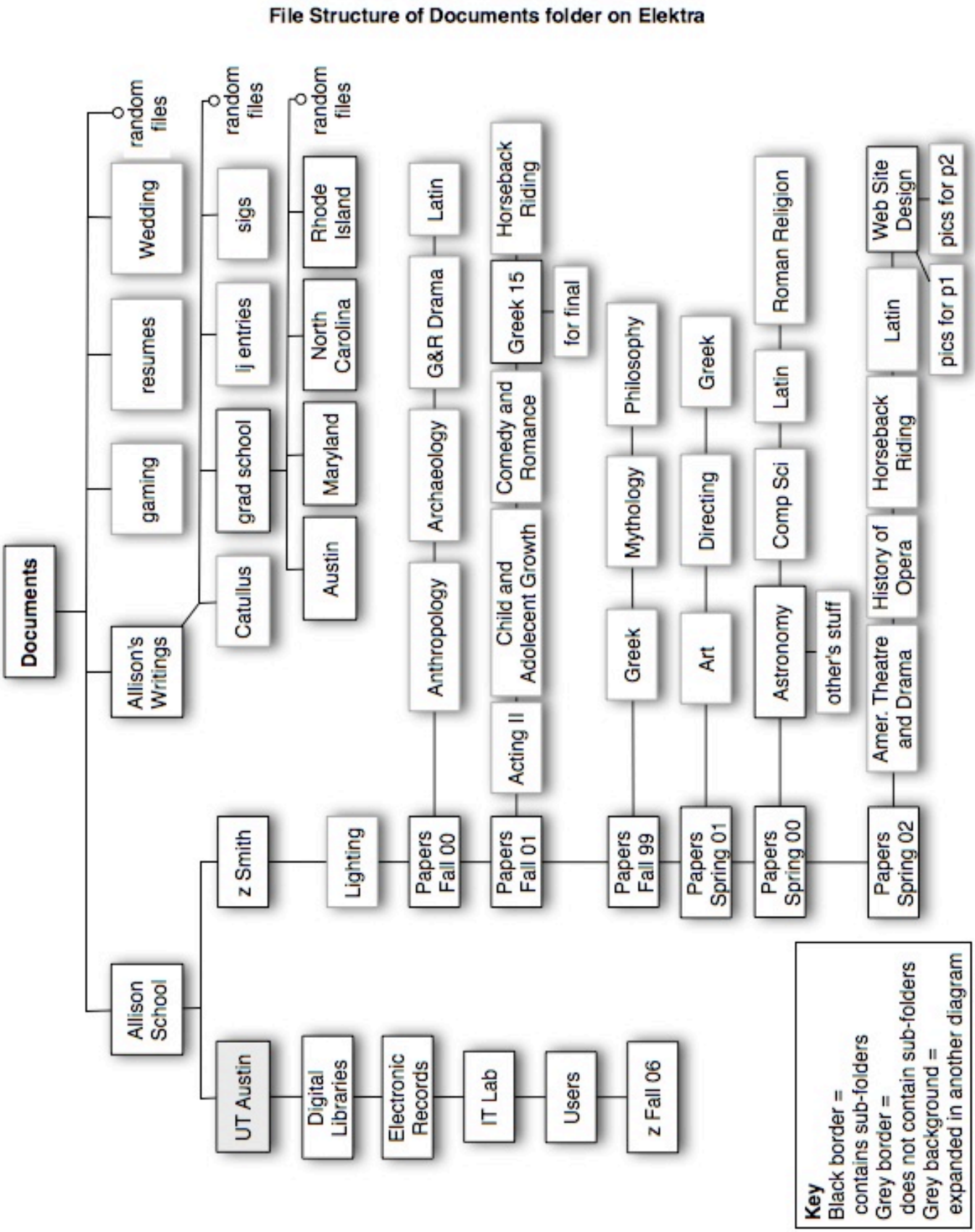
File Structure of Documents folder on Elektra
Documents
Allison's Writings
random files
Wedding
resumes
gaming
grad school
random files
sigs
lj entries
Catullus
Austin
Maryland
North Carolina
random files
Rhode Island
Allison School
UT Austin
Digital Libraries
Electronic Records
IT Lab
Users
z Fall 06
z Smith
Lighting
Papers Fall 00
Papers Fall 01
Papers Fall 99
Papers Spring 01
Papers Spring 00
Papers Spring 02
Anthropology
Archaeology
G&R Drama
Latin
Acting II
Child and Adolecent Growth
Comedy and Romance
Greek 15
for final
Horseback Riding
Greek
Mythology
Philosophy
Art
Directing
Greek
Astronomy
Comp Sci
Latin
Roman Religion
other's stuff
Amer. Theatre and Drama
History of Opera
Horseback Riding
Latin
Web Site Design
pics for p1
pics for p2
Key
Black border = contains sub-folders
Grey border = does not contain sub-folders
Grey background = expanded in another diagram

Chart 3

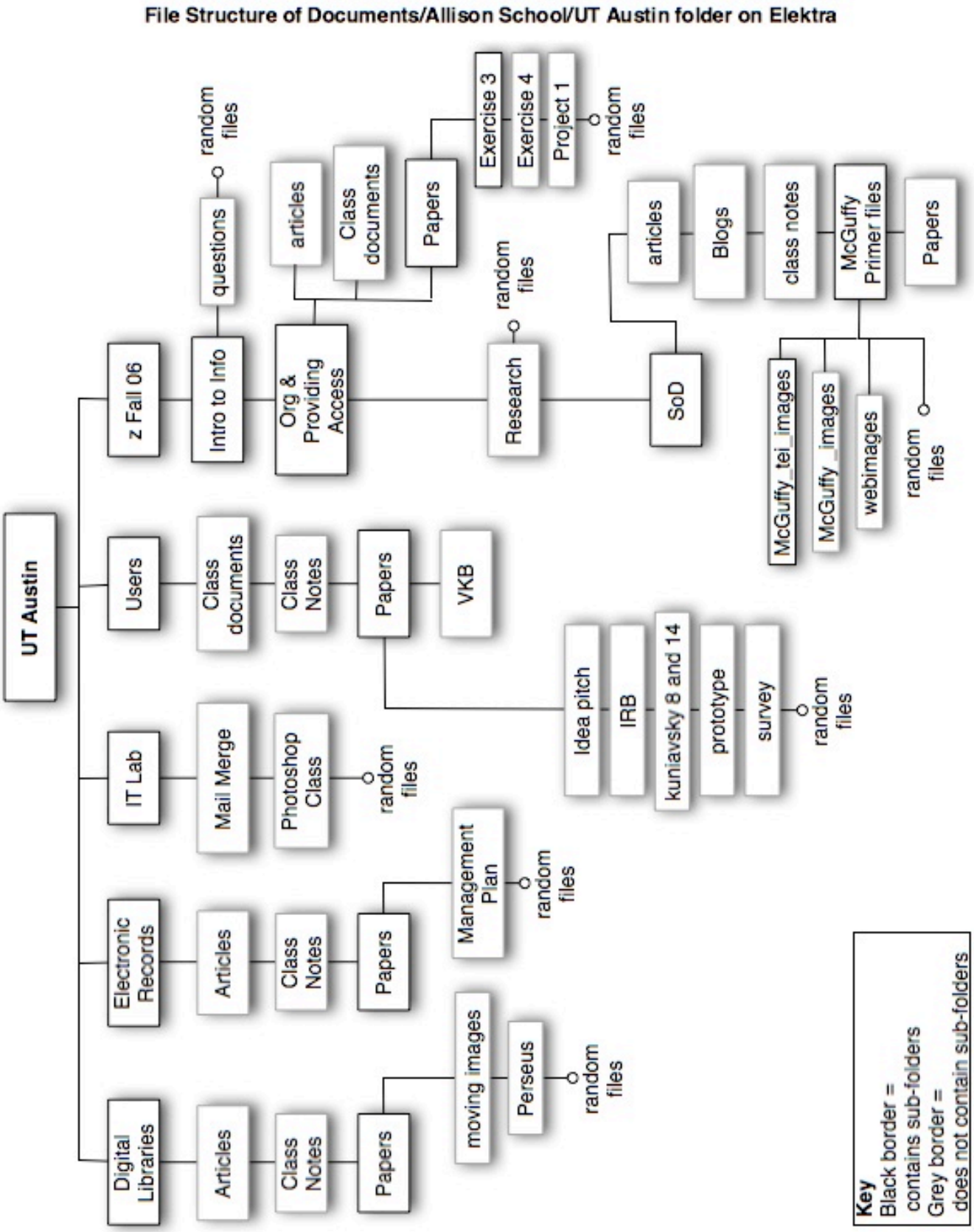
File Structure of Documents/Allison School/UT Austin folder on Elektra
UT Austin
z Fall 06
Intro to Info
questions
random files
Org & Providing Access
articles
Class documents
Papers
Exercise 3
Exercise 4
Project 1
random files
Research
random files
SoD
articles
Blogs
class notes
McGuffy Primer files
Papers
McGuffy_tei_images
McGuffy_images
webimages
random files
Users
Class documents
Class Notes
Papers
VKB
IT Lab
Mail Merge
Photoshop Class
random files
Idea pitch
IRB
kuniavsky 8 and 14
prototype
survey
random files
Electronic Records
Articles
Class Notes
Papers
Management Plan
random files
Digital Libraries
Articles
Class Notes
Papers
moving images
Perseus
random files
Key
Black border = contains sub-folders
Grey border = does not contain sub-folders

Chart 4

Number of records in each main directory of Elektra as of around April 12, 2007

File type	Desktop	Documents	Movies	Music	Pictures	Sites	Totals
.aif				1			**1**
.avi			39				**39**
.band				3			**3**
.c		1					**1**
.cdp				1			**1**
.css					2	4	**6**
.doc	1	273			1		**275**
.dtd		3					**3**
.flv			1				**1**
.gif		10			116	12	**138**
.graffle		6					**6**
.html					29	127	**156**
.idraw					3		**3**
.imovie			2				**2**
.indd		1					**1**
.itc				165			**165**
.jap					4		**4**
.jpg	550	277			1400	562	**2789**
.js						1	**1**
.m4a				18			**18**
.m4p				32			**32**
.m4v				4			**4**
.mov		4	1				**5**
.mp2				1			**1**
.mp3				1840			**1840**
.mp4			1				**1**
.mpeg			2				**2**
.pdf	6	51		5	3	3	**68**
.png					8		**8**
.ppt		25					**25**
.psd		3			11	3	**17**
.rm			2				**2**
.rtf		2					**2**
.sfk				2			**2**
.tiff		115			8		**123**
.txt		20		1	4	22	**47**
.wav				2			**2**
.wmv			3				**3**
.xjournal		13					**13**
.xls		5					**5**
.xml		5		1		1	**7**
.xsl		1					**1**
.xvkb		12					**12**
unknown		3		2			**5**
total files	**557**	**830**	**51**	**2078**	**1589**	**735**	**5840**
size (GB)	0.47	1.24	17.87	7.34	0.78	0.13	**27.84**