Fall November 29, 2017

Avoiding the Gambit For Our Personal Digital Archives

Matt Schultz, *Grand Valley State University*

Avoiding a Gambit for Our Personal Digital Archives

Matt Schultz

Introduction

In the game of chess a *gambit* move is a calculated opening salvo to wittingly sacrifice one's own assets on the board in order to draw out an opponent and gain a strategic advantage over them throughout the game. The unassuming and expendable *pawns* are often the first choice lambs, with the ultimate objective being to open up and expose an opponent's *king* to an inescapable series of advances.

The gambit is an apt analogy for the effort that today's tech giants—Apple, Google, Amazon, Facebook, and others—are investing to build online platforms and content ecosystems wedded to the advancements of the various smartphone, tablet, netbook, and wearable devices that drive users to further and further reliance upon them. In this analogy our personal digital documents, photos, audio/video, app data, and much else are the king content. Little by little our pawns of control, choice, privacy, and ownership are being drawn out, exposed, and taken down as we pipe our content more heavily into and through these devices. Indeed, with each new device purchase and "digital locker" subscription we appear to be slowly but surely migrating the full lifecycle and strata of our personal digital libraries and archives into the domain of the cloud platforms that support them.

With this paradigm shift comes a host of questions and challenges. How can we collectively uphold a culture of ownership, control, and at-will sharing of our personal archives in the face of the mounting trends of remote storage, gated access, restrictive content licensing, copyright policing, pervasive online streaming, and digital rights management (DRM)? Most importantly, how will libraries and archives contend with these ascendant platforms? What meaningful role, if any, will our public-oriented, public-minded curators play in the face of this corporately consolidated content lifecycle?

This chapter explores these questions and challenges through a series of brief and concise case studies on recent media transformations—including those involving our personal documents, our digital photos/audio/video, and our vital app data. The chapter will shed light on the various corporate- and consumer-driven market forces that are consolidating this paradigm shift. Specific threats and challenges that are posed to our personal digital archives and by extension to the born-digital acquisitions of libraries and archives will be addressed. The chapter will then conclude with a summary of practical examples of curatorial strategies and grassroots initiatives that can be employed to help mitigate an encroaching personal "Digital Dark Age."

Case Study: Personal Documents

We'll begin with personal documents—our word processor files, our spreadsheets, our presentation files. These all have a unique set of vulnerabilities when it comes to shifting our content to hosted online platforms for greater ease of access and use. Admittedly, as a class of digital materials it can be challenging to do them full justice in this conversation because the platforms that service their lifecycle are only mildly interoperable and differ from one another across a number of features. However, against the backdrop of this paradigm shift, what can be said commonly about this class of materials is that they are squarely at the center of a significant corporate push to more robustly and elegantly sync our local document edits to a remote cloud storage environment.

Syncing, as it turns out, is a strategy less designed to ensure a thorough series of archived snapshots of our documents, but rather to ensure that only our most recent edits are successfully collected, stored, and pushed to the other devices that we may be using to complete our final editing. Leading platform services like Google's Docs/Drive, Apple's iCloud/Drive, and Microsoft's Office365/OneDrive have software schemes in place to save iterative working-draft

versions of an edited document within their cloud storage environments. However, it is important to understand that these services are often then pruning these numerous snapshots down to a more limited set of stored changes that are in turn available to the document author at a later date. These limited but retained versions are the sum of what the author can return to for reference purposes and to use as a basis for restoring an older version to take the place of their most recent edits.

The dilemma, particularly for those who are editing and storing their documents solely in the cloud (e.g., Google Docs) as opposed to also syncing copies to their local "Drive" apps, is that downloading and saving any earlier versions as well as any annotations of their documents can often be somewhat fraught and time-consuming. Indeed, retrieving a full series of stored snapshots from document hosting platforms often requires enterprise level programming that utilizes Advanced Programming Interfaces (APIs). Acknowledging that not all versions are necessary or equal, the current default export features of many hosted services do not typically support download of document versions at all.

Archivists are only recently coming to terms with acquiring personal papers solely in electronic formats, and embracing the reality that acquiring working-drafts of items like manuscripts and screenplays requires a combination of digital forensics, format migration, and emulation.[1] These sorts of strategies, along with outreach/education, will be imperative for archivists going forward when it comes to acquiring cloud-supported documents. With that in mind, it would behoove archivists to familiarize themselves with the various platform-specific Drive services and:

---

[1] Dan Rockmore, *The Digital Life of Salman Rushdie*, The New Yorker, last modified July 29, 2014, http://www.newyorker.com/tech/elements/digital-life-salman-rushdie.

- Be prepared to advise donors/creators on the trade-offs for choosing and using the above-mentioned services
- Learn about the enterprise-level APIs for the various services and how they might be leveraged to facilitate more robust document stewardship and transfer across platforms
- Understand how best to configure versioning settings to ensure that local copies are reliably synced and stored
- Educate donors/creators on good local document management practices to better guarantee that valuable changes to documents are saved and distinguishable from other similar versions

When all is said and done, though documents created in proprietary cloud-oriented formats may not render the same way in their standalone app counterparts or in applications running on competing platforms, we can still be thankful at the present moment that when it comes to our personal documents these commercial cloud platforms are supporting a locally synced storage option and several interoperable export formats. This gives users a measure of assurance, confidence, and control over their creations and ensures a point of free exchange between donors and archives. The chapter will have more to say on this below.

Case Study: Personal Photos, Audio, and Video

When compared with personal documents, storage for personal digital photos and audio/video (AV) is getting steadily more complicated. For the average creator, the genesis of this sort of content has shifted rapidly to their smartphones and tablets. Users are getting all too familiar with the string of notifications from their devices alerting them to having reached their local storage limits and/or the need to purchase more remote cloud storage to continue with their

device's automatic backup routines. Device manufacturers and service providers are scrambling to offer customers flexible options for managing this conundrum locally. For numerous smartphone and tablet owners this means resorting to the previous scheme of saving born-digital photos and AV to removable microSD cards, which are unfortunately prone to being damaged and misplaced. This is also a scheme that introduces yet another step in the already challenging process of transferring content to another computer to accomplish a further chain of backup.

The other common strategy to which users are resorting is that of purchasing devices with larger storage allotments. This strategy forestalls the frequency with which users are receiving notifications about reaching their local limits, but it does very little to stave off notifications from any connected cloud storage services. The end result tends to be two-fold in that users' reliance on increased tiers of device-connected cloud backup either continues to grow, or they fall back to perpetually expanding their local storage to support transfers of content from their devices. The latter strategy, over time, hastens the rate at which we find ourselves in need of purchasing new computers with larger hard drives and creates a chaotic array of ageing external storage media that tends to collect dust on shelves.

From a consumer perspective this state of affairs feels a bit stagnant and paralyzing. Where is the elegant, convenient, end-to-end solution to this quagmire of content management?

It is early days yet for a broad uptake of capacity and skills but the archives community is making great strides when it comes to working on technical levels with collections of born-digital photos and AV materials that are donated to archives on external media. We are less prepared, however, for a future scenario wherein the paradigm flips more comprehensively towards users pushing waves of device-generated photos and AV solely to the cloud in order to avoid more frequent computer upgrades and managing stacks of dusty hard drives, SD cards,

CDs and DVDs. Storing offline copies of photos and AV materials on such media could one day be the exception as opposed to the norm. What then?

At the risk of over-speculation, it may be in archives' interest to begin investigating and testing ways to make their acquisition processes more interoperable with cloud storage platforms, either through a particular service's sharing features or through the service's programmatic APIs. This was alluded to above in the section on personal documents. The intention being to meet users halfway while keeping a hand in the content pipeline—asserting some influence and direction over how these services fit into the work of libraries and archives into the near future. Otherwise, institutional archives could very well find themselves without a stake in these commercial cloud services—all of whom are seeking to extend their current enterprise-level support and features to everyday users to help them manage the full lifecycle of this highly valued content—up to and perhaps including long-term preservation. Indeed, services such as Amazon Glacier, Google Near Line Storage, and Oracle's Archive Cloud Storage are all positioning these services as loosely defined archival endpoints within a longer chain of intermediary storage use on behalf of their customers. Partnerships between commercial giants like Google and state level archives through the Google Cultural Institute, as just one such example, should also make clear to us that these entities see themselves as playing a meaningful role in collecting and servicing digital heritage materials.

In addition to becoming more versant with the various services and APIs, further measures that archivists can undertake in this area would be to:

- Become active learners when it comes to this rapidly changing space of digital imaging and time-based media, as we simply cannot continue to approach these objects as standalone files or static content—they are highly dynamic and complex
- Educate users on how they can proactively curate their digital photos and AV materials leveraging the features and functionality of their mobile devices, social media accounts, and media libraries
- Raise more awareness about both the challenges and affordances of removable and portable storage media, the need for active media refresh schedules, and encourage responsible local and geographically distant backup procedures

Ultimately, the end goal should be to better ensure that users can continue to work easily with archives to retain control of their content and have the means to transfer their personal photos and AV for the sake of independently preserving our broader collective history. At the same time we can and should continue to ask critical questions along the way regarding the impacts to content quality and integrity as they pass through various cloud storage platforms and related sharing services. Concerns around compression and reformatting abound. Libraries and archives have to proactively engage the users, the services, and the issues.

Case Study: Personal App Data

Finally, an underserved area of focus when it comes to addressing personal digital archiving is that of app data.

We can continue to conceive of documents in the context of traditional productivity software (e.g., Microsoft office), and we can continue to conceive of photos and AV materials in the context of traditional media libraries (e.g., iTunes). But what happens as smartphone and

tablet apps begin to play a more pivotal role in our intellectual and creative workflows—by way of note-taking apps for example? Furthermore, what happens when the genesis of our personal photos is not only the default camera app on our smartphones but the filter-rich camera features bundled within a social media app? Even thinking beyond text, images, and AV materials, we realize quickly that apps have penetrated the very surface of our biology by dint of wearable health monitors and other similar haptic-driven devices and related programs.

Where does all this data ultimately live and how do we access it for the sake of longer-term archiving?

The first important thing to note is how fundamentally entangled apps are with very specific platforms, versions of operating systems, and remote databases and servers. Our smartphones, tablets, and wearable devices serve as content creation points, but they also store that content in formats highly unique to the apps that created them. These apps in turn also push our content over the internet in the form of data to be remotely stored, analyzed, (in some cases enhanced), and served back to us within the confines of the creating app or another web interface.

This of course raises all sorts of questions for a user around what rendition or elements of their data they might like to preserve and in what ways they might like to interoperate with the data outside of the native confines of their devices and associated web services? In practice users will inevitably discover that they are often wholly at the mercy of app developers and service providers. Everything boils down to *if*, and *how easy or hard*, an app service makes it to export data and in which formats. Evernote, Instagram, and FitBit, as just a few specific examples, provide several flexible options and formats for exporting user-generated data.

However, it is important to understand that the supported options and formats often dramatically restructure and re-contextualize the arrangement and organization scheme of the data itself. When attempting to access this same data outside of the app or in a separate app that might interoperate with the provided export formats, there are no sure guarantees that all of the information will survive or be easily accessible. One is at the very least promised an altogether different presentation and user experience with the newly exported/imported data.

Librarians and archivists can and should familiarize themselves with the range of different apps that users may be employing to create and manage data that falls under their desired collecting areas. A couple of ways of inventorying such apps and advising on best practices would be to:

- Survey and interview existing donors/creators about what apps and platforms they may be using
- Seek out respectable online technology and app ranking sources to cross-reference
- Compile a list of recommended apps and services, preferably ones that are cross-platform—narrowing down apps that provide flexible options for exporting user data—data that is held in open or interoperable formats
- Build out statements in collection and digital preservation policies that communicate any and all adopted support strategies for donated app datasets
- Point out preferred export formats
- Be prepared to consult with donors/creators to help them avoid using apps that seek to lock up their data

- Drive home the message that apps and app features change rapidly and sometimes disappear altogether. It is imperative that they schedule routine manual exports and backups of their most crucial app data

Perhaps most importantly, libraries and archives need to prepare themselves to grapple with challenging provenance and arrangement questions, not to mention various format migration issues when it comes to donated app data. Endnote, for example, exports user notes and attachments into folder hierarchies that do not necessarily reflect a user's model for organizing the information. In addition, the exported note formats, though somewhat open in nature (HTML, Endnote-specific XML) are not necessarily immediately amenable to reproducing the information in other formats that would approximate the original look and feel—a common challenge that archivists and digital preservationists refer to as "significant properties, "which are merely the essential characteristics of a digital object that must be preserved over time for the digital object to remain accessible and meaningful.

The Paradigm Shift in Context

These are by no means the full range of content genre and media-type transformations that archives and users are witnessing and experiencing through this paradigm shift. But they do provide a helpful starting point for understanding the forces and dynamics that are at the center of that shift.

The various tech giants are pushing more adoption of cloud platforms, cloud-dependent devices, and gated and licensed content ecosystems because it places more control in their hands for managing the complete lifecycle of content with a view towards greater market dominance and monetization. The verdict is already firmly *in* that user adoption, attention, and consumerism

are the keys to that monetization. User content and their interactions with it represent a rich source of metadata for tech giants, which enable these corporations to not only fine-tune the user experience and improve product uptake, but to also attract advertisers and publishers/distributors of content as a source of vital revenue. The more that these corporations can cordon off and control content distribution and user access/interaction the more effective it will be for them to mine that use data and leverage it to market new services, and form partnerships with advertisers and content producers.

It is not just metadata related to our Netflix views or scans of our music library via iTunes that these corporations are interested in tracking. They are interested in the metadata (and in some cases even the content) related to our documents, and our photos, our AV, our notes. All of the case studies covered above represent vital sources of information about our interests, our productivity patterns, our network of friends and collaborators, among so many other things. Bruce Schneier's 2015 book entitled *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World* explains in good detail the widespread practice of corporate (and government) data mining of user content and the role of metadata.[2] All services approach data mining differently. For example, as of the writing of this chapter Google does not mine the content of Google Docs for ad purposes, though it does in fact mine the content of Google Photos for the purposes of enhancing the app service itself.

Some of the targeted strategies employed for securing more control over the lifecycle for this content include:

---

[2] Bruce Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World, (New York: W.W. Norton & Company, Inc., 2015).*

- Leveraging premiums on local mobile device storage and marketing cheaper (sometimes free) remote cloud storage as a means of funneling users to greater use of their cloud storage platforms;
- Using digital rights management (DRM) to manage and limit content access, sharing, and portability; and
- Impeding interoperability across platforms through use of proprietary software and formats, controlled APIs, and silo-ed adoption and development of standards

All measures that by design result in fewer choices and lock-in for users, and fundamentally challenge the on-going work of archivists and preservationists.

This chapter would be remiss if it did not acknowledge the role that users themselves are playing in driving trends. For the vast majority of users the convenience and short-term cost savings that this paradigm shift represents are of significant interest. Moving to simpler, lighter-weight, portable computing, and shifting the burden of managing digital storage to the cloud is a worthwhile prospect. For many users this shift to hosted content and streaming services has opened up a whole new world of dynamic interaction and collaboration that is making the role of digital content in our lives exciting on a cultural, social, and even business level. Users cannot be faulted for taking advantage of these amazing developments in personal computing.

The fact remains however, that in many cases users may not perceive their choices as a potential and deeply profound trade-off, or as I've framed it here, a gambit. It has already been alluded to, but what users sacrifice in increasing measure for this convenience and cost-savings are not just the pawns of control, choice, and privacy, but also the pawn of outright ownership. Users may be slowly but surely ceding their rights and ability to claim ownership over their king

content and determine how or whether it can be flexibly migrated, preserved, and accessed over time, not to mention the ways in which it can or should be shared, exchanged, or retrieved from within the walled-in confines of the platforms and ecosystems to which they are being entrusted. There are a number of vectors that threaten user ownership, among them:

- Service Agreements – often containing obfuscating legal language that empowers the commercial provider to set and change the terms of use for uploaded and stored content at the expense of users' preferences
- Lack of Legal Protections – enabling commercial providers to share our content and data with other commercial and governmental entities to their own profit or for their own indemnification without our prior knowledge or consent and with very little legal recourse for protection or compensation
- Sudden Service Shut-Down – resulting in hastened opportunities for data recovery and often times the total loss and destruction of our hosted content

All of which amounts to a troubling gambit to be sure, but one that is thankfully not yet fully consolidated. Users, surprisingly enough, have a number of trends working for them, and unwitting allies in unexpected places. As just one example we should recognize that even though Apple, Google, Amazon, Facebook, and others make money from consolidating their hold over the content uploaded by individual users, they also have had their sights set on increasing their market hold over larger enterprise-level customers and even government data. All of whom bring with them stronger legal muscle for ensuring that data remains accessible, controllable, interoperable and secure from within the commercial cloud. Attempts to overly gate intellectual

content at these enterprise-levels would likely prove counter in the short term to many of the tech giants' efforts to capitalize on services to that market segment. Particularly in the area of personal documents and apps the various commercial cloud service providers have bent over backwards to ensure that content created in the cloud can be stored locally and remain portable to other formats.

Challenges & Opportunities Ahead

Libraries, archives, and other cultural memory organizations are facing a new future when it comes to working with individual creators/donors. There is no telling what role, larger or smaller, that smartphone, tablet, netbook, and wearable devices will play in relationship to other more traditional computing solutions. That future situation will ultimately be settled out or become more consolidated along the lines of some of the trends described above. It is safe to say, however, that we are already at the doorstep of an established and growing paradigm. The greatest danger is that libraries and archives will confuse this new paradigm of computer media with previous generations of fixed technology. This new class of devices is not necessarily engineered to extenuate our already ageing notions of "desktops" and "file cabinets" and "folders" and "documents," and they do not respect single points of locality or transfer. Nevertheless, our collections will only grow and acquire more richness insofar as we gain familiarity with these technologies and their users. Exploring and understanding points of interoperability and the best forms that outreach and advocacy should take are paramount.

In the best of scenarios libraries and archives will have time and presence to work closely with a donor to maneuver their hybridized or more strictly cloud-based collections into a place (or places) of networked interoperable exchange—perhaps through the use of a variety of intermediary services (e.g., DropBox, WeTransfer, Google Drive, etc.). The real struggle will be

securing the necessary financial and technology-related resources to advance testing and interoperability with the various cloud storage services that have been mentioned throughout this chapter.

Will the institution interoperate with users and their devices through only one cloud storage service, or a broad range of services? How will the institution sustain this paradigm of data acquisition and adapt to fluctuations in service subscription costs over time? What are the appropriate storage subscription tiers to handle programmatic transfers of data from donor devices? Will local computer workstations be configured with sufficient storage to download and receive any transferred data for further processing?

Along with the technical interoperability there are also questions that will need to be explored on policy levels for working within such a framework. What data management requirements will need to be placed on users when it comes to transferring their personal digital archives via such services? Will data need to be packaged or placed in certain formats to meet archive specifications, or to avoid compression and other unstated transformations that might threaten the quality and integrity of the content during transmission?

In the worst case scenario, libraries and archives will be handed this new class of devices in the absence of their owner/creator only to discover that their only access to the related data is through layers of passwords, encryption, digital rights management, and proprietary app data formats—all furthermore highly dependent upon an operating system that may be completely unsupported by the receiving computing environment.

Thankfully, the knowledge and skills that libraries and archives are beginning to amass through the acquisition of legacy computer media and the use of born-digital forensics tools/approaches can serve these new types of exchange. Through those methodologies curators

are gaining experience with troubleshooting the intricacies of port connections and peripheral devices, engaging the research required to make sense of outdated operating/file systems, and learning to leverage metadata to sift for provenance and authenticity. Libraries and archives are also amassing more and more expertise in the areas of emulation and file format migration for born-digital materials. All of these skills are helpful starting points for navigating the approaching waves of donated smartphones, tablets, netbooks, and wearable devices that we are potentially liable to receive. Not to mention the accompanying transfers of zipped-up export archives from various app services and/or extracted photo and media database files, among other similar esoteric content.

Libraries and archives are also well prepared to engage in the advocacy that will be needed to help everyday users understand their role in shifting these trends in directions that empower them on behalf of their king content. Volumes like this one are evidence of our community's passion to inform and provide guidance on issues of social and cultural importance as they pertain to technological change. This chapter itself in many ways owes its writing to inspiration from recent initiatives such as the Library Freedom Project, which is seeking to inform and equip libraries and their users with skills and technologies to protect patron privacy in our new age of surveillance.[3] The Electronic Freedom Foundation (EFF) is another reliable partner organization for libraries and archives that is integrally involved in similar efforts, as well as others such as the Apollo 1201 Project, which is geared towards eradicating digital rights management (DRM).[4] Other previous advocacy work to be built upon would be that of the Library of Congress's coordinated outreach curriculum on Personal Digital Archives, and the Digital Preservation

[3] "Our Work," *Library Freedom Project*, accessed February 16th, 2017, https://libraryfreedomproject.org/ourwork/.
[4] "About," *Electronic Frontier Foundation*, accessed February 16th, 2017, https://www.eff.org/about.

Outreach & Education (DPOE) Train-the-Trainer series.[5][6] Not to mention the work of the Digital POWRR group based at Northern Illinois University.[7] Which is all to say that librarians and archivists clearly know how to rally on behalf of users and their content!

Conclusion

It is clear then that libraries and archives are well positioned to begin working with donors/creators to preserve personal digital archives that have their genesis and significant portions of their lifecycle managed via smartphones, tablets, netbooks and wearable devices. As we continue that work in earnest there are a number of targeted strategies that we can undertake to solidify that position.

Libraries and archives must increase their familiarity with this unique class of devices, working hands on and with donors, to understand the ways in which they are becoming entangled with the online platforms and content ecosystems that support their use. More importantly, we must explore how to leverage the most open features of these devices and platforms to ensure that they have a meaningful place in our arsenal of acquisition pathways. To some degree the paradigm is set and the ship is sailing. The more that libraries and archives engage these technologies now the better position we will be in to shape them and continue curating collections into the future.

Libraries and archives should also be proactive in their outreach to donors, educating them in how best to protect their vital pawns of control, choice, privacy, and ownership in support of their king content. Likewise, libraries and archives increasingly represent a powerful market segment to tech giants and we should not undersell the role that we play in shaping the

---

[5] "Personal Archiving: Preserving Your Digital Memories," *Library of Congress*, accessed February 16th, 2017, http://www.digitalpreservation.gov/personalarchiving/.
[6] "Digital Preservation Outreach & Education," *Library of Congress*, accessed on February 16th, 2017, http://www.digitalpreservation.gov/education/index.html.
[7] "About POWRR," *Digital POWRR*, last modified on January 6th, 2016, http://digitalpowrr.niu.edu/.

development of their online platforms and content ecosystems. When it comes to concerns around the consolidation of control over our content via their market-driven gambits, libraries and archives need to engage the issues, as individual institutions but also at consortial levels, through professional associations, and via grassroots initiatives, advocating for more control, choice, privacy, and ownership. The battlefront issues of digital rights management, interoperability, and industry support for more open non-proprietary formats should be paramount in those efforts.

With this new paradigm of devices and content wedded so intricately to online platforms and content ecosystems, in some ways the challenges for libraries and archives have never been more daunting. But as adept curators we've already accumulated a knowledge base and set of expertise in areas of format migration, digital forensics, and emulation. We are ready to engage the technical challenges. It is now time to get to work raising awareness within our institutions, piloting new acquisitions, appealing for needed resources, adjusting our policies—continuing to do what we do best.

Avoiding a Gambit for Our Personal Digital Archives

Matt Schultz

Bibliography

"About." *Electronic Frontier Foundation*. Accessed February 16th, 2017. https://www.eff.org/about.

"About POWRR." *Digital POWRR*. Last modified on January 6th, 2016. http://digitalpowrr.niu.edu/.

"Digital Preservation Outreach & Education." *Library of Congress*. Accessed on February 16th, 2017. http://www.digitalpreservation.gov/education/index.html.

"Our Work." *Library Freedom Project*. Accessed February 16th, 2017. https://libraryfreedomproject.org/ourwork/.

"Personal Archiving: Preserving Your Digital Memories." *Library of Congress*. Accessed February 16th, 2017. http://www.digitalpreservation.gov/personalarchiving/.

Rockmore, Dan. "The Digital Life of Salman Rushdie," *The New Yorker*. Last modified July 29, 2014. http://www.newyorker.com/tech/elements/digital-life-salman-rushdie.

Schneier, Bruce. *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World.* New York: W.W. Norton & Company, Inc., 2015.