

The Right to Be Forgotten and Implications on Digital Collections: A Survey of ARL Member Institutions on Practice and Policy

Virginia Dressler and Cindy Kristof

In the spring of 2017, digital librarians and digital collection managers at member institutions of the Association of Research Libraries (ARL) were surveyed on practices and policies surrounding takedown requests in openly accessible digital collections. The survey collected basic demographic information surrounding the digital repositories (anonymized) and presented a series of hypothetical scenarios for respondents to consider and reflect upon. The survey received a 25.8 percent response rate, with many intriguing insights. Survey findings are presented, along with a discussion on future recommendations for work in this area.

Introduction

Digital librarians have day-to-day procedures, guidelines, and benchmarks in place for several aspects of the job, such as digitization specifications, metadata standards, and other best practice workflows. There are, conversely, many other aspects of the job that are grayer areas, more nuanced and varied with regard to decision making and common practices. This article addresses one such area: takedown requests in openly accessible digital collections. Policies and practices among digital librarians and their institutions vary, sometimes widely, in response to takedown notices and, as such, frequently base their decision making on situational context, sometimes in consultation with institutional counsel.

A survey conducted in the spring of 2017 focused on the process of decision making by digital librarians for takedown requests in digital collections at ARL member institutions. Whereas the Right to Be Forgotten (RTBF) court case (*Google Spain SL, Google Inc. v Agencia Española de Protección de Datos, Mario Costeja González*) in 2014 was centered on a takedown request by a private citizen to a major search engine of an old news item printed a decade before, our survey focuses on requests made directly to the institution publishing the digital content. As a consequence of the 2014 court case, Google now has a method in place to handle such requests (43.3% removal rate

Virginia Dressler is Digital Projects Librarian and Assistant Professor, and Cindy Kristof is Head, Copyright and Document Services, and Associate Professor, both at Kent State University; email: vdressle@kent.edu, ckristof@kent.edu. ©2018 Virginia Dressler and Cindy Kristof, Attribution-NonCommercial (<http://creativecommons.org/licenses/by-nc/4.0/>) CC BY-NC.

of 873,873 URLs reviewed to date, as of January 13, 2018, provided on the overview page in the “Google Transparency Report”¹), and a clear process in place to address and weigh requests.² Further, Google has also publicly provided the outline of the takedown process and provides examples of valid requests for the purpose of reference.

The survey presented a number of hypothetical scenarios to digital librarians, asking them to consider how they would acknowledge and process these requests. Demographic information was collected on the respective digital collections and provided digital librarians the opportunity to share any real-life anecdotes in the same vein. In the paper, the major themes that emerged from the study are outlined, as well as the role the presence (or absence) of policy may play within these institutions. The overall goal of the survey and findings were to create a dialogue around the processes revolving around takedown requests and perhaps serve as a catalyst toward further research into, or development of, standards and practices.

Background

The notions of privacy and the right to privacy are certainly not new concepts in human society. In 1890, Samuel Warren and Louis Brandeis published “The Right to Privacy”³ in the *Harvard Law Review*, navigating issues of identity, publication, and individual rights. The authors articulated and encouraged the concept of the “right to be let alone” as a right of an individual to enjoy life through privacy in an increasingly more documented and public world. Warren and Brandeis’ article addressed the implications of printed media (text and photographs) to define these new rights for individuals and likened them to the right not to be assaulted or beaten, defamed, or maliciously prosecuted. The authors asserted the necessity of defining new rights to meet the newer demands from society as a consequence of recent technological innovations.

Even further back in the history books, during the 15th and 16th centuries in Italy, there was an opportunity for newly released prisoners, on the occasion of three significant annual festivals, to destroy any and all record of past offenses under the supervision and blessing of the presiding Duke in Ferrera.⁴ Reflecting on this example, Gabriella Giannachi stated in *Archive Everything*, “Problematic past actions or histories would at that point be erased and new lives could be started,”⁵ in so providing a mechanism for individuals to essentially start anew. While these particular examples are drawn from life before computers and networks, it is interesting to reflect upon the timelessness of human behavior and the concept of individuals’ rights to privacy, along with the desire for a degree of power over one’s past persona.

Bert-Jaap Koops from the Tilburg Law School wrote of the internet having an “iron memory”⁶ with access to mountains of data on any given topic and how pieces of data could be categorized as either digital footprints (data created by the individual) or data shadows (data generated about individuals by others). This is an intriguing concept to apply here in this debate and gets to the question of who is publishing the information and by what rights does one have this information. Koops specifically mentions issues revolving around bankruptcy law, juvenile criminal law, and credit reporting. These are of particular interest in the topic of the RTBF and the rights of an individual to this information, especially when one considers the notion of forgiveness within each of these examples. The author also brings a more philosophical perspective in the discussion of the concept of forgetting, something that may be a truly human component that does not exist naturally in the digital world.

Likewise, in the fascinating title *Delete* by Viktor Mayer-Schönberger,⁷ the notion of the ability for one to forget in order to begin anew is discussed as crucial to our own history and humanity. Suggesting doing this programmatically through an expiration date of sorts, Mayer-Schönberger calls for an equivalent to forgetting to take place in

the world of infallible machines and constant accessibility to information of all sorts. There are many areas of debate on this topic to explore, all of which are still active across a number of disciplines such as law and philosophy. On the idea of having an infinite online archive and user selection of expiration dates of content, Koops stated, "you never know when you might want to see it back in the future,"⁸ which is an intriguing comment on the fickleness of human nature. The overall ethical implications of decisions regarding takedown notices aside, this article focuses more specifically on works with clearer publishing boundaries than some of the more broadly defined "data" examples above.

Although the underlying principles of privacy and the need to make a fresh start in one's life are at the crux of this issue and are certainly not new concepts, the aspect that sets these needs apart in modern society is, of course, the internet: the ubiquitous availability and easy access to information. Where does this issue of one's right to privacy and "forgetting" stand with regard to openly accessible digital collections? Can policy provide a framework for digital librarians and institutions looking for clarification? Does an individual have a right to privacy within digital collections (broadly defined here), or are the First Amendment rights of free expression more important? Do individuals deserve privacy and forgetting more than the public deserves an accurate and unabridged account of events? Can copyright guide us in some cases? These are some of the concepts raised within this survey.

Literature Review

The current literature in the United States on the RTBF is scattered in the popular press, the fields of law, technology, and communications, the library and information science journals related to law, technology, and communications, and in journals that focus on archival content.

Most articles reviewed for this study discussed the "Google Spain" case (Google Spain v. González, 2014 E.C.R. 317), arguably the most famous of its kind to date and the "first test of the RTBF law in the European Union."⁹ This case was the result of a 2010 complaint filed with the Agencia Española de Protección de Datos (AEPD), Spain's data protection agency, by Spanish citizen Mario Costeja González. Using Google, González had found links to a story published in 1998 by the Spanish newspaper, La Vanguardia Ediciones SL, about an auction of real estate holdings he had held to pay his social security debt. Arguing that his financial problems were resolved, part of his past, and no longer relevant, González requested that La Vanguardia remove the identifying information surrounding this story and that Google Spain and Google, Inc. remove all links in its search results. González's request of the La Vanguardia was rejected by the AEPD on grounds that the story was published by order of the Spanish Government; however, the AEPD asserted it had the authority to order Google to put search blocks in place.¹⁰

After an appeal by Google to the Spanish high court, the Audiencia Nacional, questions were referred by the latter to the European Court of Justice (ECJ), which concluded in May 2014 that, under the European Union law, Google was responsible for removing such search engine links.¹¹ Google subsequently created a form allowing Europeans to request to have search results of names removed, allowing requesters "to list one or more URLs they want delisted," which applies to Google's European domains only.¹² The "RTBF ruling does not affect the original published content... only the search engine results," and "does not apply to the U.S. site google.com."¹³

Of course, the attention to the "Google Spain" case has brought the opposite of privacy to Mr. González. This is known as the Streisand Effect, coined in 2005, for singer Barbra Streisand's¹⁴ attempts to remove from circulation photographs of her personal residence, which further brought more attention to the original content. Though

Google's European procedures keep requesters for information removal anonymous, Xue et al.¹⁵ were able to identify 80 requesters by analyzing the data that Google does share. A content analysis of delisted articles revealed topics such as sexual assault, murder, financial misconduct, pedophilia, and terrorism. As noted earlier, there is around a 43 percent removal rate of submitted takedown requests.

The literature consulted describes in great detail the current differences in American versus European jurisprudence regarding the natural tension between the individual's right to privacy and the public's right to free expression/access to information. Although both aspects exist in both sets of laws, the former tends to dominate in European law while the latter takes precedence in American law. The European Union's General Data Protection Directive contains "the idea of erasure"¹⁶ while the United States treasures its First Amendment. Robert Larson¹⁷ argues that Europe's privacy interests, which arise from the French "droit à l'oubli" (roughly translated to "right to oblivion"), are ultimately "incompatible with free speech." There are exceptions, of course; Chris Conley of the ACLU of Northern California¹⁸ argues for the rights of the individual to deletion and privacy. A.M. Klingenberg of the University of Groningen in the Netherlands discusses the "public interest not to forget" as in the Dutch Act on Archives,¹⁹ "freedom of expression" as in the case of *Times Newspapers v. United Kingdom*, and the "right of access to data" as outlined in Article 8 of the Charter of fundamental rights of the European Union. A 2017 experiment (Bode and Jones) on American attitudes toward the RTBF revealed that the concept is more likely to be supported if "either websites or search engines are in charge of execution" (as opposed to a government agency), if the RTBF focuses on the rights of children, if the RTBF excludes criminal information, and if there is no limit with regard to the age of the information itself.²⁰

From a European point of view, Pekka Henttonen thoroughly discusses at length the problem of digital archives, records management, and privacy,²¹ describing five strategies for approaching the problem of protecting privacy. The first of these is called the "purpose limitation principle," in which personal information is collected if it is compatible with purposes specified when collecting the data. This idea is put to use internationally by four privacy instruments. The second approach is "privacy self management," in which individuals are given power over their personal information, either regarding its usage or its deletion/destruction (in other words, RTBF). Problems that surround the former involve a lack of understanding of privacy policies and issues keeping people from making good choices. The RTBF is, in Europe, contained within the right of the personality, which includes "dignity, honor, and the right to a private life." A third idea, outright destruction of data by decision of a third party, involves waiting for a time period to pass and retaining the data only if it is still needed for its original purpose. In this sense, destruction becomes an extension of the purpose limitation principle; however, the individual's choice and involvement in this process are not retained. Anonymization is the fourth idea, offered as a compromise solution; the data is retained and can be used but is no longer traceable to individuals. However, this solution can reduce the usefulness of the data, and attackers with malicious intent may be able to reattach data to identities. The final strategy is the information safe haven approach in which information is held within the custody of an archive and released only at an appropriate time to protect individuals during the course of their lives. The author notes that, with the exception of privacy self-management and the RTBF, most of these methods are "paternalistic" in nature, as they involve decisions made by a third-party authority and not by the individual. The author recognizes the enormity and difficulty of managing personal data in a digital world.

Survey Results & Methodology

The survey was sent to ARL institutions between February and March 2017. At the time of the survey, 124 institutions in the United States and Canada were included. Appropriate staff from each institution were identified to which to send the survey, namely those whose job titles seemed to indicate direct involvement in managing digital collections. The survey was provided in both English and French. The response rate was 25.8 percent, with just 2.4 percent of the responses discarded for being incomplete. The full list of the questions, as well as the answers to the quantitative questions, are provided in appendix A. The survey was sent to the identified individuals through Kent State University’s Qualtrics institutional account after review and approval from the campus’s Institutional Review Board (IRB), #17-059.

Results

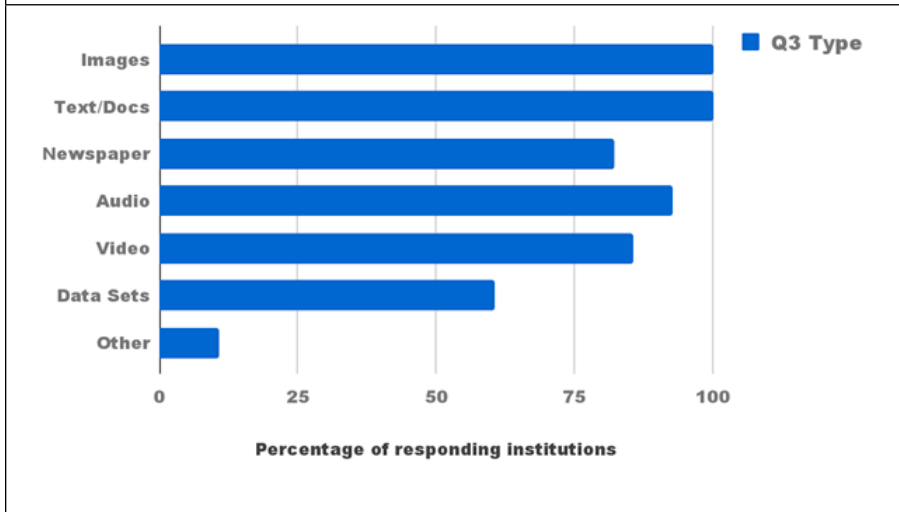
Demographics of Digital Collections: Questions 1–9

All responding institutions indicated that they have digital collections of some nature, with most of the institutions (93%) having had digital collections online for more than five years. More than one-fifth (21%) of institutions indicated having had online collections for 11–14 years, while 39 percent indicated having had online collections for more than 15 years.

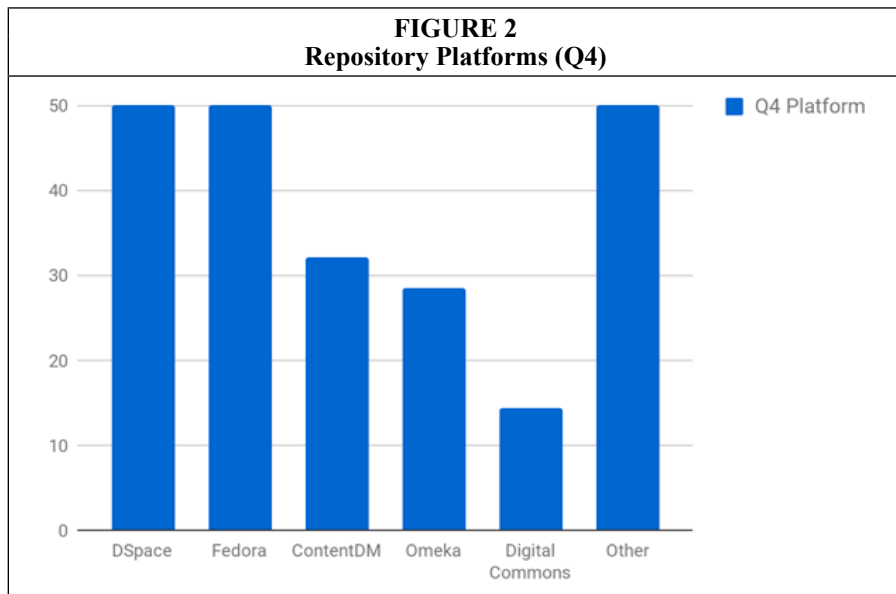
All of the digital collections surveyed contain both text and images, with more than 75 percent also containing newspaper, audio, and video. More than half of digital collections store data sets as well. In the “Other” category, respondents included e-books, student and faculty scholarship, and journals, which one could argue belong in Text/Documents category with the examples provided in the parentheses.

With regard to overarching themes in digital collections, most survey respondents indicated a large emphasis on archival collections, with regional history following up as the second most popular type of content. Faculty research topped student research by about 25 percent. Respondents also indicated research collections and state-level specialized collections. Most of the respondents (23) indicated they host some kind of newspaper collection or content.

FIGURE 1
Types of Digital Content in Surveyed Institutions’ Digital Repository (Q3)



Respondents indicated a wide variety of storage and access platforms for their digital collections. Repositories backed by DSpace and Fedora were the most prevalent. A number of institutions specified other or locally designed solutions, which included: YouTube, Scribd, Archive-It, Artstor, HathiTrust, Kaltura, Mukurtu, DLXS, WordPress, Open Journal System, Luna, and Internet Archive.

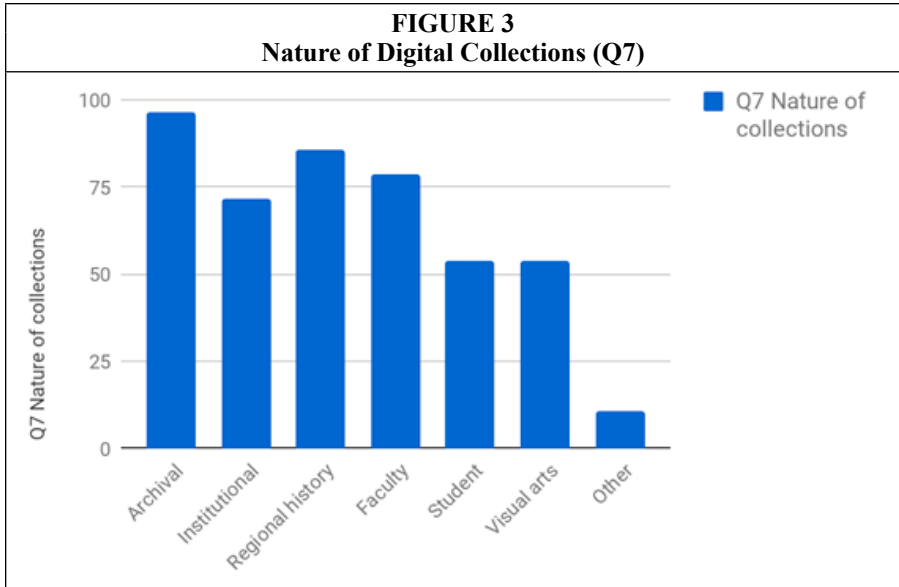


Regarding public accessibility, 13 institutions indicated there are restrictions to access for some parts of their digital collections. These restrictions are mainly due to issues surrounding copyright (musical performances, restrictions from the author, embargoes, and the like). None of the surveyed institutions indicated that the collections were completely closed or inaccessible.

The varying sizes of ARL institutions are reflected in the number of staff involved in some aspect of digital collections. All but two respondents have at least two full-time positions in place to address digital projects, though many note that this work is in addition to other job responsibilities. The average across the ARL institutions is 2–3 full-time equivalents working on digital projects, with the largest reporting nine full-time employees dedicated to digital projects. Some of the job titles reported include: Digital Projects Librarian, Digital Initiatives Coordinator, Open Access Repository Coordinator, Digital Repository Librarian, and Digital Scholarship Services Librarian (figure 3).

There are some variances, too, in the size of repositories by the approximate number of items/objects. The majority of the respondents (61%) indicated there are more than 50,000 objects in their digital repositories. Only one respondent was unsure of the total number of items (which, for institutions with multiple platforms in place, this can be a difficult question to answer quickly in a short survey). Nine institutions (32%) responded they have 10,000 to 49,999 objects, and the remaining respondent related that they have fewer than 5,000 objects.

Institutions were then asked if they had a policy, procedure, or guidelines in place to address takedown requests of content on their website. This question was intentionally broad to apply not only to digital collections but also to other aspects



of the institutional website. Eleven institutions (40%) responded that they have a policy in place, with seven providing either a link to, or an uploaded copy of, the policy. This question will be discussed more in depth later in the article, with a policy analysis. Nine institutions (32%) responded that they do not have any kind of policy in place, and eight institutions (29%) reported a draft is in the works to address this issue. One of the more interesting responses received in this question was “Yes—we have a commitment to academic freedom—we will take down content that violates copyright, but won’t with a knee jerk reaction take down content that someone finds offensive.”

Hypothetical Scenarios

The next part of the survey revolved around a set of hypothetical questions, which were intentionally left open-ended to allow for the survey participant to interpret notions of privacy, takedown requests, and internal processes. The responses were open-text responses with no limitation on length.

You receive a request for a name to be removed from a particular item in your digital library, directly from the individual in question. The requester claims that the inclusion of their name in an openly accessible digital library violates their privacy. The name appears in print in your digital regional newspaper collection, within the student newspaper that was published in print at your institution and later digitized for the digital collection. This content has been run through optical character recognition (OCR) software, and has been fully indexed by search engines such as Google. How would you respond?

Question 10 marked the beginning of the hypothetical scenarios in the survey. Some deidentified, selected responses are below. The first hypothetical question posed to the respondents asks them to think about a takedown request from the individual in question in their student newspaper collection. Free-text responses ran the gamut from a simple “I don’t know” to more thoughtful, in-depth responses.

One respondent: "I would check with our Scholarly Communications lawyer, but would assume that no change would be required—we are merely providing access to an already existing item and would not want to modify the historical record." This statement provides an insight into a process in place for such requests, and an awareness of the overall nature of the request. Another respondent said they would not take down an item unless there was a threat to the person's safety; they also presented a series of questions in place at their institution to check to see if the item is available elsewhere, ascertain the age of the material in question, and discover whether the item is considered an institutional record.

Another thoughtful response: "Administratively, keeping track of what's been redacted has been troublesome." The respondent goes on to state that such a request could spark some much needed internal discussion to set a precedent on how to handle similar requests. The same respondent also indicated such a request would require discussion with three other staff members to resolve the issue, from the Associate Dean to University Archivist to Head of the Institutional Repository; they also state that they may consult with the general counsel office as well. Again, there was at least an awareness of the individuals who should be involved in such a discussion and an acknowledgment of the need for consistency in response in lieu of having a policy in place.

One respondent made a distinction by publication status, stating that they will not remove any previously published material (but would, on the other hand, consider a request for an item not previously published). This reflection points to a unique characteristic of digital collections: they frequently contain a blend of published and unpublished content. This may be a pivotal component in some decisions, particularly when an individual's rights are weighed into a takedown request.

Three respondents stated that they would immediately remove access to the item in question while the issue is being resolved. If the takedown request is approved, one institution stated they would then obscure only the page(s) in question if possible (rather than the whole issue or volume). Another institution from this subset stated, "We would redact the name somehow if the person felt strongly about it." And the third stated, "We would maintain the digital representation of the newspaper while removing the name from the OCR text file to prevent crawlers from indexing the name and making it easily discovered." Two others responded with a hypothetical action of possibly removing the name, though they would do so only after heavy consideration internally, with one noting that they "... may remove the personal name from the metadata but would not remove the newspaper from the online collection."

One respondent did not need any further information than the initial request to make the decision in favor of the requestor to completely remove the item from its digital collections and move into a dark archive, then further stated they would work with the appropriate staff from Information Services or a respective department to remove the specific information from search results.

Four respondents indicated that their general counsel would be consulted on any takedown request, and two respondents indicated an exact time frame they would provide the requestor to address the concern. Two respondents also offered that they would use this request as a way to educate the person on copyright issues and the rights of the university to publish content in its digital collections: "The student newspaper was a publicly available document, and copyright is owned by the University." One respondent stated that they have a copyright policy in place that would address many concerns but are working on a privacy policy currently. Another respondent simply replied, "Don't waste my time—Find yourself a lawyer and schedule an appointment with them to talk about your privacy," a particularly surprising, blunt, and ultimately unhelpful response to find within a service-oriented profession.

And finally, one respondent stated, "We would discuss their reasons and explain it's a news source and we can't change it. It would be unethical to alter news from the past. If they claim the article is defamatory, we would refer them to University Council." Clearly, there can be a marked difference between theoretical considerations and the perceived ideals surrounding privacy when compared with real-world responses and practices.

You receive another request to remove a name from another digital object from the digital newspaper collection. In this scenario, you find that there is a later mention of a correction to a story that could aid in the requester's defense. (Misprinted information, subsequent findings that alter the original story, a court case where the person is later found innocent of charges, and so on.) This particular newspaper was not published by your institution, but from a local township. How would you respond?

The next hypothetical question dealt with the idea of a misprint in the original newspaper that was later corrected to the requestor's defense. This subtle change in the question was intended to get the survey respondents to think about the idea of a printed document as potentially fallible.

Five respondents said this scenario did not change anything in their minds when compared to the first hypothetical question, and they would proceed in the same manner as the original request, which would be to retain the current version and not alter or remove information.

One respondent, while sympathetic, stated they would leave the article as is unless there was a safety concern but added that they would try to add a link to the correction. Another sympathetic respondent replied, "We would make an effort to create a link to the correction, or provide a reference to the correction, but would not go to extraordinary lengths."

Another respondent said they would first consult with the publisher of the newspaper to get advice on this situation. One respondent said they would pose the question to their advisory board and library leadership to determine what to do, indicating a desire for consensus from a larger group of people to address the takedown request.

One respondent delved a bit further on this question, stating that it depended on if the newspaper issue in question was born digital or not. Though this factor was not specified in the second question, it is an intriguing notion in terms of whether this is, or should be, a factor.

You receive a request from a publisher to remove an article from your institutional repository. The individual has published a portion of an article that was originally published in your institutional repository. The publisher is threatening a lawsuit and has also requested that you have the results removed from search engine results. How do you respond?

The next hypothetical question is focused on a takedown request from a publisher regarding an article in an institutional repository, for an article that has been published in part elsewhere. Nine institutions responded that, if the request is proven to be a valid one, they would work to remove the article and attempt to work on removing as much information from search engine results as far as they could: "We would then contact Google to ask that they clear their cache for that particular URL." Several respondents spoke to the different role and rights associated with institutional repositories as contrasted with other digital collections. A few of the respondents would ask the publisher directly to provide proof of the agreement with the author, while some would take the opportunity to try to contact the author directly to clarify the situation.

One respondent indicated that they have a strategy in place "... to deal with the grey area of who published what first." Eight respondents said they would involve general counsel immediately since there is a threat of a lawsuit against the institution, with one adding an honest quip that they would "secretly want to tell [the] publisher to take a long walk off a short pier but aim for diplomatic detachment."

In the scenarios listed above, could you please think about the person(s) who would be charged with these decisions at your institution? Below, list their job titles. Include other relevant information as well; such as if a committee or working group is in place to address these types of issues, or if such requests would be directed to library administration to address.

A side interest in this survey was to address who would be charged with such decisions as outlined in the above questions, and also if larger working groups or committees were in place to address such requests. Eight respondents included a Scholarly Communications Librarian as being a key person to consult, and other titles/departments included: Head of Digital Initiatives, University Archivist, General Counsel, Library Dean, Copyright Librarian, Digital Librarian, various Associate Deans, one advisory board, and one library leadership team. Many cited a method such as a regular monthly meeting with key individuals to address these types of scenarios, with a small minority stating these decisions could be made by one or two individuals on the fly.

Finally, if you have had a real-life scenario that is similar to the ones listed above, could you provide information below illustrating such a scenario? Please describe the request, the subsequent chain of events internally, persons involved in the resolution, and the outcome.

Many of the responses to these real-life scenarios were quite interesting, as they revealed the variations in how requests are handled and who is charged with the decision making.

One respondent said, "We have many such requests including the removal of culturally sensitive material, removal of proprietary company information (included inadvertently in student dissertations), removal of improperly published content. Copyright ownership and cultural sensitivity are our two major decision points." Another respondent stated that they have received requests from students wanting to revise a thesis or dissertation, to which they will allow the addition of an errata sheet for corrections.

One respondent replied that their institution has redacted personal information such as Social Security numbers in the past.

Another respondent provided a scenario of hosting public records, including how they would supplement a public record if requested but not alter it. In a real-life case of a person with a negative aspect in their public record, they denied the request to remove the record. Two respondents cited cultural sensitivity as being a big factor when weighing questions of privacy in digital collections, with one respondent citing the *Protocols for Native American Archival Materials* to provide some guidance for requests. Another common request for one respondent was the removal of information contained in wedding announcements and arrest reports, wherein the digital repository has been advised to work directly with the publishing unit who will then advise on the removal (which to date has not resulted in the removal of any content). Likewise, another respondent would advise requesters to seek legal guidance to prompt any removal, also wherein there have been no removals to date.

One of the more interesting personal anecdotes was the removal of an oral history file, in which the interviewee had mentioned a fact about another individual that the

individual's family felt was slanderous. Since the institution was courting the latter's family for a potential donor, the request was permitted, and access to the file is provided only upon request.

Another institution described the digitization of their theses and dissertations en masse and will take down a file upon direct request of the alumnus.

One institution described the steps to take due diligence to address copyright and privacy concerns and noted, "these steps help, but don't of course completely prevent any future scenarios [as described in this survey]."

Analysis of Results

Demographic Questions

The information provided in the first two questions provided a framework for the survey, in that most of the responding ARL institutions have larger, robust, and well-established digital collections that are quite varied in scope. One could surmise that more established digital repositories would be more likely to have a solid framework in place for takedown requests, as such requests are likely to crop up over time.

Further, many institutions reported using more than one platform. One peripheral takeaway is that no single platform currently available serves all digital collections needs within the modern research institution. It was also interesting to see the variety of responses in the write-in area of "other" in question 4. These entries reflecting a wide variety of digital service points, from external services such as YouTube to Flickr, to internally maintained servers and digital repositories.

Role of Policy

The libraries surveyed were largely from the United States, in which the First Amendment takes priority and ALA's Library Bill of Rights echoes this concept, where respondents would almost always hesitate to hinder or remove access to materials. As such, it was the thought of this research team that these institutions would err on the side of preserving the historical record. The variety of responses and the rationale behind them were a surprise, revealing a lack of uniformity in contemporary practice and with regard to notions of privacy.

Respondents were asked to either provide a link to an existing policy or policies or directly upload the document into the Qualtrics survey. Seven respondents (25%) were able to provide some type of policy that may relate to takedown requests in digital repositories. In many of these examples outlined below, only a small amount of information within the policy actually addresses the issue. In some policies, the process of a takedown request is briefly described, but most do not specify much detail regarding the overall procedures or involvement of specific library or institutional personnel. For the purpose of maintaining anonymity, the authors have removed specific references to any identifying pieces of information. Six of the respondents with policies are academic institutions, and one is a public library. Of the six academic institutions, half of these are private institutions.

Within the submitted policies, there was variation in the depth and breadth of the implications of action. Two of the institutions provided policies that refer more to overarching information access and security and that only peripherally address takedown processes. Issues such as online conduct and expectations of staff and student behavior are emphasized, with one heavily citing the Digital Copyright Millennium Act (DCMA) to cover online collections and issues pertaining to copyright. Only three policies provided specific detail on how one would go about requesting takedown, contact information, and the projected turnaround time to address requests.

One of the most helpful policies outlined hypothetical scenarios where items may be considered for withdrawal from the repository, mainly due to copyright concerns. If an item is removed from the repository, the policy details the action of the library creating a “tombstone” page to alert users that the item is no longer accessible and also mentions that the item will be kept in a separate, inaccessible archive after removal from the public-facing site. Another institution states that an item will be removed from public view until a takedown request is resolved, which was interesting as it is not clear how long the review of the request might take.

One public-facing takedown policy examined in this article (and cited by one submitted policy) is from HathiTrust.²² This policy echoes many of the elements already addressed above and provides users with a concise page that provides a mechanism to contact the repository, outlines required information for requests, and gives an expected turnaround time on decisions. This policy could serve either as a model or as the basis for beginning discussions on policy issues for institutions that currently do not have procedures in place.

The issue of whether or not all institutions should have specific takedown policies is perhaps for another study to examine, though it is the feeling of the authors that personnel responsible for digital repositories should at least consider and discuss the issues surrounding takedown requests and the RTBF. For many, policies may provide a framework for forethought regarding their collections; but, for others, takedown requests may be better handled on a case-by-case basis, given all the potential variables. A working group, committee, or other set of appropriate personnel named to this task may be desirable so as not to leave such requests on the shoulders of one individual.

Hypothetical and Real-World Scenario Answers

The answers provided in the hypothetical portion of the survey were a fascinating insight into the varying outlooks of ARL practitioners, with answers ranging from unsure and indicating inaction, to referral to another party, to the complete removal of information upon request. However, the answers provided do point to a need in the profession to create a framework for evaluating such requests, as well as the need for more education and discussion on the issues of publication, privacy, and ethics of data removal. The wide variety of current takedown practices and related policies (formal or informal) may have negative implications to the historical record of the future, and the notion of a true, fully indexed, searchable digital archive becomes muddied. At times, decisions regarding takedown result in direct action taken to edit or alter a part of an openly accessible digital collection and/or related metadata record. This has lasting, and oftentimes hidden, consequences on the notion of a truly open and fully indexed digital collection.

As the answers in the real-world scenarios demonstrate, there may at times be a need to remove information for various reasons. With this in mind, there is a need to communicate the possibility of data removal occurrences to the end user, who may assume that all content is searchable and indexed. Will future researchers be aware of the variety of practices among institutions? Will they know they must, or will they be willing to, dig deeper for hidden history? Perhaps some combination of strategies as described by Henttonen in the literature review above can be used to achieve a balance. For example, some takedown requests might be moved to a dark archive with an expiration date.

Role of Legal/General Counsel in the Institution

The issue of how the general counsel office at an institution (or comparable office) could potentially be involved in the scenarios described in the survey is more of a

peripheral area of the article, though (as many respondents highlighted) the office can be a crucial part of decision making and response to such requests. Relationships likely vary widely between general counsel offices and the various academic libraries and other institutions surveyed, though there is the potential to leverage a relationship between these two units that can improve both services by understanding what the other does in a more meaningful way. Further, general counsel offices can also be of great assistance when potential or actual liabilities arise, as well as drafting new policies and procedures. These offices also often serve as a legal protection when policies become practice.

Recommendations for Further Research and Takeaways for Practitioners

As indicated in the survey results, there is a lack of standard policies and practices; it illuminated enormous differences in how takedown requests are currently handled. However, a recent, related lawsuit can be referred to for some guidance on the topic. Although the legal system itself is not a flawless one and can be influenced by a number of outside factors, recent litigation may nevertheless prove to be quite helpful in illuminating procedures for institutions. We have seen examples of this with copyright cases. Copyright expert Kevin Smith noted, "Because copyright has not kept up with the changes in technology, court cases are the way we learn what is or is not permissible."²³ One could say that the RTBF notion as applied to digital collections may follow suit.

In 2008, an alumnus of Cornell University filed a lawsuit after the library digitized and shared the weekly newspaper broadly online through its digital repository, eCommons. The student claimed that the university had libeled him and disseminated private information online. The case was ultimately dismissed, citing that the article in question was truthful and accurate and therefore the university was not intent on defaming the individual. U.S. District Court Judge Barry Moskowitz further noted, "Truth is an absolute defense to any libel action."²⁴ Cornell University Librarian Anne Kenney stated that the dismissal of the court case supported the idea of creating a comprehensive digital archive and ultimately also supported the notion of making documentary material more accessible. Further, Kenney stated, "I do share concerns that individuals might have about potentially embarrassing material being made public, but I don't think you can go back and distort the public record."²⁵

Since the 2014 RTBF decision, there have been some further developments. In July 2017, the European Court of Justice was set to rule on a case between Google and the French government whether or not the RTBF could extend beyond the European Union borders.²⁶ These new developments could either work to reverse the initial ruling or make further strides to define this notion. Stateside, in February 2017, a bill (A05323) to create an RTBF Act was introduced in the New York State Assembly. It sought "to rectify damaged reputations of individuals whose lives have been affected through inaccurate information found online."²⁷ At press time, the bill had been "referred to governmental operations" and no floor votes had taken place. No matter the outcomes, the ramifications of case law and legislation are bound to shape the RTBF as the years pass, and it is only a matter of time before this happens.

Conclusion

At a bird's eye view, the answers provided in this survey proved that, to date, there are no clear answers for digital librarians in the real world and often a lack of clearly defined practices in regard to takedown requests. Often, the response to an individual's takedown requests come from one practitioner, reflecting their point of view, and not that of a clearly defined policy or practice. The professional librarians surveyed displayed a broad array of personal opinion and thought processes revolving around takedown

requests, and this is evident in the outcomes of such requests. Institutions may not wish to or be able to conform fully to professional organizations' model policies or standards of practice once they are developed; but, at the very least, discussions should take place across institutions and patrons should be provided with contact information.

There is additionally a larger looming question that this survey did not address: what are the overarching implications of these variable current practices on a fully searchable digital collection? For example, the institutions that may remove an item or mention of the name for any requester, regardless of the merit of the request, may create a ripple effect within the larger information structure. This could have a deleterious effect on the value and definition of open access digital collections, creating a "swiss-cheese" effect, with black holes of removed data in a digital archive. The longer-term implications of such varying practices are immense, and often not blatantly apparent to the end user. As a profession, such scenarios need to be addressed with a high level of fairly applied principles and rationale. Library and information professionals are the architects of future history, and this record deserves discussion, thought, and care.

APPENDIX A. Right to Be Forgotten and Digital Collections Survey

Thank you for your participation in our survey. The survey should take between 10 and 15 minutes to complete, depending on the time spent answering the more qualitative, free-text questions. The Institutional Review Board (IRB) at Kent State University has approved this study. Your participation is completely voluntary, and by clicking "I Agree," you are also confirming that you are at least 18 years of age. The answers provided in this survey will be collected anonymously, and we have ensured the software will not collect any personal or identifying information. There will first be a series of multiple-choice questions that will address some basic demographic information regarding digital collections at your institution, followed by some hypothetical scenarios that we would like you to consider regarding privacy and digital collections. You may also optionally provide relevant links or documents in addition to these survey answers. The information we gather in the survey will be used to write an article, and any identifying information given in the free-text answers will be anonymized. The investigators listed below will be the only persons with direct access to the information provided by participants through the survey software, Qualtrics.

It is our hope that participants will provide thoughtful, truthful responses to the questions posed in the survey. In the case that additional information is provided by the participant, we will guarantee that any identifying information of the institution will be removed in the event of inclusion in a publication and will further be securely deleted at the end of the project.

If you are 18 years of age or older, have understood the statements above, and freely consent to participate in the study, click on the "I Agree" button to begin the experiment. Please feel free to contact one of the investigators listed below if any additional information is needed or any other concerns arise. We very much appreciate your time and we value your input.

Q1 Does your institution have digital collections?

Digital collections here are defined broadly as any unique, unlicensed online resources that are provided by your institution.

- Yes
- No

Responses

(28) Yes 100%

Q2 Approximately how long have your digital collections been online?

- 0–4 years
- 5–10 years
- 11–14 years
- 15 years+
- Unknown

Responses

(1) 0–4 years 3.57%
 (9) 5–10 years 32.14%
 (6) 11–14 years 21.43%
 (11) 15 years+ 39.29%
 (1) Unknown 3.57%

Q3 Select the types of digital items in the digital library. Select all that may apply.

The Right to Be Forgotten and Implications on Digital Collections 987

- Images
- Text/Documents (include books, journals, archival content, and the like)
- Newspaper
- Audio
- Video
- Data Sets
- Other (please specify) _____

Responses

(28) Images 100%

(28) Text/Documents (include books, journals, archival content, and the like) 100%

(23) Newspaper 82.14%

(26) Audio 92.86%

(22) Video 85.71%

(15) Data Sets 60.71%

(3) Other (please specify) (e-books, student and faculty scholarship, journals) 10.71%

Q4 What platforms are your digital collections hosted on? Select all that may apply.

- Fedora/Hydra
- DSpace
- ContentDM
- Digital Commons
- Omeka
- Other/Locally Designed Solution (Please specify details, if available)
- Don't Know/Unsure

Responses

(14) DSpace 50%

(14) Fedora/Hydra/Islandora 50%

(9) ContentDM 32.14%

(8) Omeka 28.57%

(4) Digital Commons 14.29%

(18) Other/Locally Designed Solution 64.29%

Locally designed platform, YouTube, Flickr, Scribd, Archive-It, Artstor, HathiTrust, Home grown, Django/Python, Solr, ImageMagick, Mukurtu, Drupal based websites, Kaltura, Blacklight on Solr index, homegrown php system, DLXS, WordPress, Open Journal System, Silverstripe website CMS, Luna, Open Journal Systems, Internet Archive, internal servers

Q5 Are your digital collections openly accessible?

- Yes
- No
- Partially (please explain) _____
- Don't Know/Not Sure

Responses

(15) Yes 53.57%

(13) Partially 46.43%

Q6 What is the current staffing for digital projects in place at your institution? Please describe the number of positions, job titles, and if the position is part-time or full-time.

Q7 What is the general nature of your digital collections? Check all that may apply.

- Regional/Local History

- Institutional History
- Archival Collections
- Visual Arts
- Faculty Scholarship
- Student Scholarship
- Other (Please describe) _____

Responses

- (27) Archival Collections 96.43%
- (20) Institutional History 71.43%
- (24) Regional History 85.71%
- (22) Faculty Scholarship 78.57%
- (15) Student Scholarship 53.57%
- (15) Visual Arts 53.57%
- (3) Other (Please describe) 10.71%

Medical slides, research collections, more so than faculty scholarship, specialized collections

Q8 Could you provide a rough estimate of the size of your digital collection, counting each unique digital item as one object.

- 1–999 objects
- 1,000–4,999 objects
- 5,000–9,999 objects
- 10,000–49,999 objects
- 50,000+ objects
- Don't Know/Unsure

Responses

- (1) 1,000–4,999 objects 3.57%
- (9) 10,000–49,999 objects 32.14%
- (17) 50,000+ objects 60.71%
- (1) Don't Know/Unsure 3.57%

Q9 Do you have a policy (or policies) in place that address/es takedown requests of content in your digital library or website?

- Yes. If available, please provide a link to the policy in text box below, or upload directly in the next question. _____
- No
- Draft in the Works
- Don't Know/Unsure

Responses

- (11) Yes 39.29%
- (9) No 32.14%
- (8) Draft in the Works 28.57%

If available, please upload takedown policy. (If there are multiple files, please either contact the researchers via email or combine into one document.)

The next questions will pose some hypothetical scenarios, and we would like you to think about how your institution would respond to each situation. In your answer, please provide additional details on how you would respond, who at your institution would be consulted (provide job titles and not names), and how you think the situation would be resolved. And it is fine if you are not sure how these issues would be resolved; please feel free to indicate this as well. Any information provided in the scenarios will be anonymized for any identifying aspects in the final publication.

The Right to Be Forgotten and Implications on Digital Collections 989

Q10 You receive a request for a name to be removed from a particular item in your digital library, directly from the individual in question. The requester claims that the inclusion of their name in an openly accessible digital library violates their privacy. The name appears in print in your digital regional newspaper collection, within the student newspaper that was published in print at your institution, and later digitized for the digital collection. This content has been run through optical character recognition (OCR) software and has been fully indexed by search engines such as Google. How would you respond?

Q11 You receive another request to remove a name from another digital object from the digital newspaper collection. In this scenario, you find that there is a later mention of a correction to a story that could aid in the requester's defense. (Misprinted information, subsequent findings that alter the original story, a court case where the person is later found innocent of charges, and so on). This particular newspaper was not published by your institution, but from a local township. How would you respond?

Q12 You receive a request from a publisher to remove an article from your institutional repository. The individual has published a portion of an article that was originally published in your institutional repository. The publisher is threatening a lawsuit and has also requested that you have the results removed from search engine results. How do you respond?

Q13 In the scenarios listed above, could you please think about the person(s) who would be charged with these decisions at your institution? Below, list their job title/s. Include other relevant information as well, such as if a committee or working group is in place to address these types of issues or if such requests would be directed to library administration to address.

Q14 Finally, if you have had a real-life scenario that is similar to the ones listed above, could you provide information below illustrating such a scenario? Please describe the request, the subsequent chain of events internally, persons involved in the resolution, and the outcome.

Notes

1. "Transparency Report," *Google*, available online at <https://transparencyreport.google.com/> [accessed 20 July 2017].
2. "European Privacy Requests for Search Removals," *Google*, available online at <https://www.google.com/transparencyreport/removals/europeprivacy/> [accessed 5 August 2017]. Information on the process and implementation of content removal can be found here: <https://docs.google.com/file/d/0B8syaa6SSfiT0EwRUFyOENqR3M/preview>.
3. Samuel Warren and Louis Brandeis, "The Right to Privacy," *Harvard Law Review* 4, (Dec. 15, 1890): 193-220.
4. Richard Brown, "Death of a Renaissance Record-Keeper: The Murder of Tomasso da Tortona in Ferrara, 1385," *Archivaria* 44 (1997): 21, available online at www.archivaria.ca/archivar/index.php/archivaria/article/view/12195 [accessed 1 August 2017].
5. Gabriella Giannachi, *Archive Everything: Mapping the Everyday* (Cambridge, Mass.: MIT Press, 2016), 5.
6. Bert-Jaap Koops, "Forgetting Footprints, Shunning Shadows: A Critical Analysis of the 'Right to Be Forgotten' in Big Data Practice" (December 20, 2011). *SCRIPTed*, Vol. 8, No. 3, pp. 229-256, 2011; doi.org/10.2139/ssrn.1986719 .
7. Victor Mayer-Schönberger, *Delete* (Princeton, N.J.: Princeton University Press, 2011).
8. Koops, "Forgetting Footprints, Shunning Shadows," 14.
9. Jane Kirtley, "Misguided in Principle and Unworkable in Practice: It Is Time to Discard the

Reporters Committee Doctrine of Practical Obscurity (and Its Evil Twin, the Right to Be Forgotten), *Communication and Law Policy* 20, no. 2 (2015): 103.

10. Kyo Ho Youm and Ahran Park, "The 'Right to Be Forgotten' in European Union Law Data Protection Balanced with Free Speech?" *Journalism and Mass Communication Quarterly* 93, no. 2 (2016).

11. Sylvia de Mars and Patrick O'Callaghan, "Privacy and Search Engines: Forgetting or Contextualizing?" *Journal of Law and Society* 43 (2016): 257–84, doi:10.1111/j.1467-6478.2016.00751.x.

12. Youm and Park. "The 'Right to Be Forgotten'."

13. Minhui Xue, Gabriel Magno, Evandro Cunha, Virgilio Almeida, and Keith Ross, "The Right to Be Forgotten in the Media: A Data-Driven Study," *Proceedings on Privacy Enhancing Technologies* 2016 4 (2016): 1–14.

14. Antoon De Baets, "A Historian's View on the Right to Be Forgotten," *International Review of Law, Computers & Technology* 30, no. 1/2 (2016): 57–66.

15. Xue et al., "The Right to be Forgotten in the Media."

16. Muge Fazlioglu, "Forget Me Not: The Clash of the Right to Be Forgotten and Freedom of Expression on the Internet," *International Data Privacy Law* 3, no. 3 (2013): 149–57, doi:10.1093/idpl/ipt010.

17. Robert Larson, "Forgetting the First Amendment: How Obscurity-Based Privacy and a Right," *Communication Law and Policy* 18, no. 1 (2013).

18. Chris Conley, "The Right to Delete," *AAAI Spring Symposium Series* (2010): 53–58.

19. A.M. Klingenberg, "Catches to the Right to Be Forgotten, Looking from an Administrative Law Perspective to Data Processing by Public Authorities," *International Review of Law, Computers & Technology* 30, no. 1/2 (2016).

20. Leticia Bode and Meg Leta Jones, "Ready to Forget: American Attitudes toward the Right to Be Forgotten," *Information Society: An International Journal* 33, no. 2 (2017): 76–85, doi:10.1080/01972243.2016.1271071.

21. Pekka Henttonen, "Privacy as an Archival Problem and a Solution," *Archival Science: International Journal of Recorded Information* 17 (2017): 285–303, doi:10.1007/s10502-017-9277-0.

22. HathiTrust Takedown policy, available online at https://www.hathitrust.org/take_down_policy [accessed 12 June 2017].

23. Kevin Smith and Susan Davis, "Copyright in a Digital Age: Conflict, Risk and Reward," *Serials Librarian* 64 (2013): 57–66, doi:10.1080/0361526X.2013.759875.

24. Bill Steele, "Libel Lawsuit over 1982 Chronicle News Item Is Dismissed," *Cornell Chronicle* (June 9, 2008), available online at <http://news.cornell.edu/stories/2008/06/libel-lawsuit-against-cornell-over-1983-news-item-dismissed> [accessed 1 June 2017].

25. Ellen Marsh, "Libel Lawsuit against Cornell University Library Digitization Project Dismissed," Cornell University Library website, available online at <https://www.library.cornell.edu/about/news/press-releases/libel-lawsuit-against-cornell-university-library-digitization-project> [accessed 1 June 2017].

26. Alex Hern, "EJC to Rule on Whether 'Right to Be Forgotten' Can Stretch beyond EU," *The Guardian* (New York, N.Y.), July 20, 2017, available online at <https://www.theguardian.com/technology/2017/jul/20/ecj-ruling-google-right-to-be-forgotten-beyond-eu-france-data-removed> [accessed 1 August 2017].

27. New York (State) Legislature (Assembly), *An act to amend the civil rights law and the civil practice law and rules, in relation to creating the right to be forgotten act*, 300-A (S 5323), 2017–2018 Reg. Sess. (Feb. 8, 2017), *New York State Assembly*, available online at http://nyassembly.gov/leg/?default_fld=&leg_video=&bn=A05323&term=&Summary=Y&Actions=Y&Committee%26nbspVotes=Y&Floor%26nbspVotes=Y&Memo=Y&Text=Y&LFIN=Y [accessed 31 July 2017].