Framing Privacy in Digital Collections with Ethical Decision Making

Synthesis Lectures on Information Concepts, Retrieval, and Services

Editor
Gary Marchionini, *University of North Carolina at Chapel Hill*

Synthesis Lectures on Information Concepts, Retrieval, and Services publishes short books on topics pertaining to information science and applications of technology to information discovery, production, distribution, and management. Potential topics include: data models, indexing theory and algorithms, classification, information architecture, information economics, privacy and identity, scholarly communication, bibliometrics and webometrics, personal information management, human information behavior, digital libraries, archives and preservation, cultural informatics, information retrieval evaluation, data fusion, relevance feedback, recommendation systems, question answering, natural language processing for retrieval, text summarization, multimedia retrieval, multilingual retrieval, and exploratory search.

Framing Privacy in Digital Collections with Ethical Decision Making
Virginia Dressler

Mobile Search Behaviors: An In-depth Analysis Based on Contexts, APPs, and Devices
Dan Wu and Shaobo Liang

Images in Social Media: Categorization and Organization of Images and Their Collections
Susanne Ørnager and Haakon Lund

Exploring Context in Information Behavior: Seeker, Situation, Surroundings, and Shared Identities
Naresh Kumar Agarwal

Researching Serendipity in Digital Information Environments
Lori McCay-Peet and Elaine G. Toms

Social Monitoring for Public Health
Michael J. Paul and Mark Dredze

Digital Libraries for Cultural Heritage: Development, Outcomes, and Challenges from European Perspectives
Tatjana Aparac-Jelušić

iRODS Primer 2: Integrated Rule-Oriented Data System
Hao Xu, Terrell Russell, Jason Coposky, Arcot Rajasekar, Reagan Moore, Antoine de Torcy, Michael Wan, Wayne Shroeder, and Sheau-Yen Chen

Information Architecture: The Design and Integration of Information Spaces, Second Edition
Wei Ding, Xia Lin, and Michael Zarro

Fuzzy Information Retrieval
Donald H. Kraft and Erin Colvin

Quantifying Research Integrity
Michael Seadle

Incidental Exposure to Online News
Borchuluun Yadamsuren and Sanda Erdelez

Web Indicators for Research Evaluation: A Practical Guide
Michael Thelwall

Trustworthy Policies for Distributed Repositories
Reagan W. Moore, Hao Xu, Mike Conway, Arcot Rajasekar, Jon Crabtree, and Helen Tibbo

The Notion of Relevance in Information Science: Everybody knows what relevance is. But, what is it really?
Tefko Saracevic

Dynamic Information Retrieval Modeling
Grace Hui Yang, Marc Sloan, and Jun Wang

Learning from Multiple Social Networks
Liqiang Nie, Xuemeng Song, and Tat-Seng Chua

Scholarly Collaboration on the Academic Social Web
Daqing He and Wei Jeng

Scalability Challenges in Web Search Engines
B. Barla Cambazoglu and Ricardo Baeza-Yates

Social Informatics Evolving
Pnina Fichman, Madelyn R. Sanfilippo, and Howard Rosenbaum

On the Efficient Determination of Most Near Neighbors: Horseshoes, Hand Grenades, Web Search and Other Situations When Close Is Close Enough, Second Edition
Mark S. Manasse

Building a Better World with Our Information: The Future of Personal Information Management, Part 3
William Jones

Click Models for Web Search
Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke

Information Communication
Feicheng Ma

Social Media and Library Services
Lorraine Mon

Analysis and Visualization of Citation Networks
Dangzhi Zhao and Andreas Strotmann

The Taxobook: Applications, Implementation, and Integration in Search, Part 3
Marjorie M. K. Hlava

The Taxobook: Principles and Practices of Building Taxonomies, Part 2
Marjorie M. K. Hlava

Measuring User Engagement
Mounia Lalmas, Heather O'Brien, and Elad Yom-Tov

The Taxobook: History, Theories, and Concepts of *Knowledge Organization*, Part 1
Marjorie M. K. Hlava

Children's Internet Search: Using Roles to Understand Children's Search Behavior
Elizabeth Foss and Allison Druin

Digital Library Technologies: Complex Objects, Annotation, Ontologies, Classification, Extraction, and Security
Edward A. Fox and Ricardo da Silva Torres

Digital Libraries Applications: CBIR, Education, Social Networks, eScience/Simulation, and GIS
Edward A. Fox and Jonathan P. Leidig

Information and Human Values
Kenneth R. Fleischmann

Multiculturalism and Information and Communication Technology
Pnina Fichman and Madelyn R. Sanfilippo

Transforming Technologies to Manage Our Information: The Future of Personal Information Management, Part II
William Jones

Designing for Digital Reading
Jennifer Pearson, George Buchanan, and Harold Thimbleby

Information Retrieval Models: Foundations and Relationships
Thomas Roelleke

Key Issues Regarding Digital Libraries: Evaluation and Integration
Rao Shen, Marcos Andre Goncalves, and Edward A. Fox

Visual Information Retrieval Using Java and LIRE
Mathias Lux and Oge Marques

On the Efficient Determination of Most Near Neighbors: Horseshoes, Hand Grenades, Web Search and Other Situations When Close is Close Enough
Mark S. Manasse

The Answer Machine
Susan E. Feldman

Theoretical Foundations for Digital Libraries: The 5S (Societies, Scenarios, Spaces, Structures, Streams) Approach
Edward A. Fox, Marcos André Gonçalves, and Rao Shen

The Future of Personal Information Management, Part I: Our Information, Always and Forever
William Jones

Search User Interface Design
Max L. Wilson

Information Retrieval Evaluation
Donna Harman

Knowledge Management (KM) Processes in Organizations: Theoretical Foundations and Practice
Claire R. McInerney and Michael E. D. Koenig

Search-Based Applications: At the Confluence of Search and Database Technologies
Gregory Grefenstette and Laura Wilber

Information Concepts: From Books to Cyberspace Identities
Gary Marchionini

Estimating the Query Difficulty for Information Retrieval
David Carmel and Elad Yom-Tov

iRODS Primer: Integrated Rule-Oriented Data System
Arcot Rajasekar, Reagan Moore, Chien-Yi Hou, Christopher A. Lee, Richard Marciano, Antoine de Torcy, Michael Wan, Wayne Schroeder, Sheau-Yen Chen, Lucas Gilbert, Paul Tooby, and Bing Zhu

Collaborative Web Search: Who, What, Where, When, and Why
Meredith Ringel Morris and Jaime Teevan

Multimedia Information Retrieval
Stefan Rüger

Online Multiplayer Games
William Sims Bainbridge

Information Architecture: The Design and Integration of Information Spaces
Wei Ding and Xia Lin

Reading and Writing the Electronic Book
Catherine C. Marshall

Hypermedia Genes: An Evolutionary Perspective on Concepts, Models, and Architectures
Nuno M. Guimarães and Luís M. Carrico

Understanding User-Web Interactions via Web Analytics
Bernard J. (Jim) Jansen

XML Retrieval
Mounia Lalmas

Faceted Search
Daniel Tunkelang

Introduction to Webometrics: Quantitative Web Research for the Social Sciences
Michael Thelwall

Exploratory Search: Beyond the Query-Response Paradigm
Ryen W. White and Resa A. Roth

New Concepts in Digital Reference
R. David Lankes

Automated Metadata in Multimedia Information Systems: Creation, Refinement, Use in Surrogates, and Evaluation
Michael G. Christel

© Springer Nature Switzerland AG 2022

Reprint of original edition © Morgan & Claypool 2018

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Framing Privacy in Digital Collections with Ethical Decision Making
Virginia Dressler

ISBN: 978-3-031-01188-7 Paperback
ISBN: 978-3-031-02316-3 ebook
ISBN: 978-3-031-00223-6 Hardcover

DOI 10.1007/978-3-031-02316-3

A Publication in the Springer series
SYNTHESIS LECTURES ON INFORMATION CONCEPTS, RETRIEVAL, AND SERVICES, #64

Series Editor: Gary Marchionini, University of North Carolina, Chapel Hill

Series ISSN: 1947-945X Print 1947-9468 Electronic

Framing Privacy in Digital Collections with Ethical Decision Making

Virginia Dressler
Kent State University

SYNTHESIS LECTURES ON INFORMATION CONCEPTS, RETRIEVAL, AND SERVICES #64

ABSTRACT

As digital collections continue to grow, the underlying technologies to serve up content also continue to expand and develop. As such, new challenges are presented which continue to test ethical ideologies in everyday environs of the practitioner. There are currently no solid guidelines or overarching codes of ethics to address such issues. The digitization of modern archival collections, in particular, presents interesting conundrums when factors of privacy are weighed and reviewed in both small and mass digitization initiatives. Ethical decision making needs to be present at the onset of project planning in digital projects of all sizes, and we also need to identify the role and responsibility of the practitioner to make more virtuous decisions on behalf of those with no voice or awareness of potential privacy breaches.

In this book, notions of what constitutes private information are discussed, as is the potential presence of such information in both analog and digital collections. This book lays groundwork to introduce the topic of privacy within digital collections by providing some examples from documented real-world scenarios and making recommendations for future research.

A discussion of the notion privacy as concept will be included, as well as some historical perspective (with perhaps one the most cited work on this topic, for example, Warren and Brandeis' "Right to Privacy," 1890). Concepts from the The Right to Be Forgotten case in 2014 (*Google Spain SL, Google Inc. v Agencia Española de Protección de Datos, Mario Costeja González*) are discussed as to how some lessons may be drawn from the response in Europe and also how European data privacy laws have been applied. The European ideologies are contrasted with the Right to Free Speech in the First Amendment in the U.S., highlighting the complexities in setting guidelines and practices revolving around privacy issues when applied to real life scenarios. Two ethical theories are explored: Consequentialism and Deontological. Finally, ethical decision making models will also be applied to our framework of digital collections. Three case studies are presented to illustrate how privacy can be defined within digital collections in some real-world examples.

KEYWORDS

privacy, ethical decision making, digital collections, digital libraries, redaction, digitization

Contents

Preface . . . . . **xiii**

Acknowledgments . . . . . **xv**

Introduction . . . . . **xvii**

1 Framing Privacy within Digital Collections **1**
1.1 Digital Collections 1
1.2 Basic Concepts 4
1.3 Privacy: Main Concepts and Overview 6
1.4 Violation of Privacy 10
1.5 Types of Personal Data 14
1.6 Defining Personal Data 20
1.7 Concealment/Unconcealment 21
1.8 A Discord between "Information (almost) Wants to Be Free" and "Not All Information Wants to Be Free" 23

2 Core Ethical Theories and Decision Making Frameworks **27**
2.1 Background 27
2.2 Consequentialism 27
2.3 Deontological 28
2.4 Ethical Decision Making: Models and Resources 30
2.5 Examples of Code of Ethics 34
2.6 Conclusion 36

3 Role of the Practitioner as Active Agent and Notions of Privacy in Digital Collections **39**
3.1 Addressing the Notion of Harm 40
3.2 The Right to Be Forgotten and the EU Data Privacy Laws 41
3.3 Conflict of Free Speech and Privacy 45
3.4 Real-World Examples: Lawsuits, Case Studies, and Other Relevant Research 46
3.5 Other Considerations for Decision Making 48

4 Core Values and Considerations for the Practitioner **61**

4.1 Ethical Decision Making and the Digital Librarian 61
4.2 Role of Policy 63
4.3 Law and the Role of General Counsel 63
4.4 Dissemination as Disclosure 64
4.5 Moving Toward Privacy Nirvana? Tavani's Perfect Privacy Theories 65
4.6 Privacy Review in Real-World Applications 65
4.7 Setting the Stage for Good Practice 67
4.8 Ethical Decision Making: Gauging Harm and Acknowledging Responsibility 68
4.9 Privacy Assessment and Review 69
4.10 Future Directions and Conclusion: Morals, Transparency, and Participation 70

Appendix **73**

Bibliography **75**

Author Biography **85**

Preface

While standards and benchmarks have been determined and established for many of the more practical aspects of a digital librarian's day-to-day existence (such as basic digitization capture standards and defined metadata schemas), there are other more nuanced and complex situations that arise to test an area of the profession, which to date has proved to be largely undocumented and under researched. In this book, varying notions of what constitutes private information will be discussed, as will the potential presence of such information in both analog and digital collections. There are currently no solid guidelines or overarching codes of ethics to address such issues. Publishing private information can have damaging consequences as a result of widely disseminating this data via open digital collections. The practitioner becomes the responsible party for this disclosure through the action of sharing digital content as part of their daily work.

Once content is published onto an openly accessible platform, there are ramifications to consider as a result of the action of publication. The responsibility of this action needs to be addressed, as well as considering selected, existing professional ethical standards that can help guide better, more ethical decision making. This book will strive to lay some groundwork to introduce the topic of privacy within digital collections by providing some examples from documented real-world scenarios and making recommendations for future research. I will also include some viewpoints that speak to my particular experiences with regard to the work we do in digital libraries, through a few personal anecdotes.

An initial discussion of the notion of privacy as concept will be included, and will include some historical perspectives (with perhaps one the most cited work on this topic, for example, Warren and Brandeis' "Right to Privacy" (1890)). Concepts from the The Right to Be Forgotten case in 2014 (*Google Spain SL, Google Inc. v Agencia Española de Protección de Datos, Mario Costeja González*) will be elaborated on as to how some lessons may be drawn from the response in Europe and also how European data privacy laws have been applied. The European ideologies will be contrasted with the Right to Free Speech in the First Amendment in the U.S., and the complexities in setting guidelines and practices revolving around privacy issues when applied to real life scenarios will be highlighted. Two ethical theories will then be briefly explored: Consequentialism and Deontological. Finally, ethical decision making models will also be applied to our framework of digital collections.

Three case studies are presented to illustrate how privacy can be defined within digital collections in some real-world examples. First, the libel lawsuit against Cornell University in 2008 will be explored in the first case study, as one of only first lawsuits to go to court around a digital collection. Then, an examination of Reveal Digital's decision to digitize and publish the full run of *On*

Our Backs, a lesbian erotica magazine which was in publication from 1984–2006 in the company's *Independent Voices* project for a second case study. The online publication of roughly 20 years of the magazine essentially outed hundreds of women by pushing a limited-run print journal to a much broader audience. Finally, a look at a less traditional "digital collection" of images within a subreddit will be explored and considered.

Finally, a recommendation of some core values and practices to consider for digital librarians will be made to provide a usable framework in decision making and project planning. While the American Library Association (ALA) has an existing Code of Ethics in place for the profession at large, the notion of privacy in the code is more centered around a patron's right to privacy, and does not address some of the other aspects of privacy and personal information that can be relevant in digital collections. Initial recommendations will be provided as a starting place for digital librarians, modeled in part from charters and principles from related writings and resources.

As digital collections continue to grow, the underlying technologies to serve up content also continue to expand and develop. As such, new challenges are presented which continue to test ethical ideologies in everyday environs of the practitioner. The digitization of modern archival collections, in particular, presents interesting conundrums when factors of privacy are weighed and reviewed in both small and mass digitization initiatives. Ethical decision making needs to be present at the onset of project planning in digital projects of all sizes, and we also need to identify the role and responsibility of the practitioner to make more virtuous decisions on behalf of those with no voice or awareness of potential privacy breaches as result of these decisions.

Acknowledgments

This book was very much inspired by one particular real-life conundrum, and the realization that even within a small working group of librarians with a collected 50+ years of varied professional experience, we can often consider notions of privacy, and one's right to privacy, in very, very different ways. This book also reflects many conversations over the last year with numerous individuals in my life who have entertained and inspired my quest into the grayer areas of digital librarianship. I feel we are overdue as practitioners to think more critically of the impact of our decisions to share, describe, and provide access to materials that may have potential privacy hazards.

I am very much indebted to be in a work environment that affords creative thinking and also offers research leave to pursue this particular rabbit hole. I am also fortunate to have an ever-supportive spouse who overlooks my Sunday morning ritual of burrowing away into a corner of the house to steal some rare, quiet moments in the midst of the chaos and insanity that comes along with raising toddler twins; to read about ethics, privacy, and notions of time and being in the digital world. Thank you, my love—this book could not have happened without your support in so many ways, by way of your constant selflessness and sacrifice.

Special thanks and shout outs to the following individuals in no particular order: Cindy Kristof, an exceptionally keen copyright librarian who serves as a frequent afternoon coffee companion and often talks me through the crossroads; Laurence Skirvin, who gave me the courage to pencil out the first inklings of this project one fall afternoon at the student union, and also serves as my favorite philosophical devil's advocate (even if he is not as fond of Martin Heidegger as I am); Jeff Rubin, my first boss in the digital projects realm to whom I owe so much of my day-to-day practice, and who took a chance on a young woman with a no real tangible job experience at the Louisiana State Museum in 2005; Michael Kavulic, who inadvertently inspired a major line of inquiry and the underlying thesis of this book, by way of his often sage and pensive reflections that I have very much come to enjoy; Doug Kubinski, perhaps the best-humored lawyer I've come to know this past year, who advised on everything from book contracts to left-field questions of legality and ethics within the digital library landscape; and last but not least, Kay Wise, grandmother extradoniare and one of the kindest souls who inspires me daily.

In addition, this book was also very much inspired by the research, writing, and lectures of the following individuals; Tara Robertson, who has beautifully articulated the ethical problems present in the *On Our Backs* digitization project; Daniel Solove, whose writings about privacy in the digital age are phenomenal, and my only wish is that I would have had more time to ponder and digest his writings more fully in the limted six months that I spent writing and researching for this book;

and Gene Pendleton, professor emeritus in the Department of Philosophy at Kent State University, who is unlikely aware of the impact his teaching and lectures had on me as an undergraduate art history major through his Aesthetics and Postmodern philosophy courses I took many, many years ago. As I worked toward the deadline to complete this title, I realized I would have enjoyed another year to read further into the mountain of books on privacy that remained unread on my desk, which only means that my fascination and captivation continues.

I am also grateful for feedback and conversations with the following folks, again in no particular order other than perhaps the recency of our conversation over coffee, lunch, drinks, or a phone call, when they may have unknowingly given me a particular thought or inspiration: Karen Gracy, Kiersten Latham, Peter Lisius, Kara Robinson, Haley Antell, Cara Gilgenbach, Thomas W. Dressler, Thomas R. Dressler, Lizette Barton, Jodi Kearns, and Jennine Vlach. And to the Chesterland Arabica and Tree City Cafe, where I often holed up on several frigid Northeast Ohio winter days, and who provided copious amounts of coffee to fuel this writing.

I dedicate this book to our twins, Winston and Gillian, as my views on an individual's right to privacy (and the desire for privacy) in the modern world have surely been impacted since their arrival into this crazy world in the spring of 2016. And also to our first child, whom I lost inexplicably at the end of a full-term pregnancy just a year and a day before the twins' arrival. He is ever present in my heart and mind, and I believe nothing in this world will teach you more about life and love than losing an infant, and will also irrevocably alter your entire being from that moment forward. Every day that has since passed, my heart aches without you in my life, little one.

Introduction

As someone who has worked in the digital library landscape for the past decade or so, I have come to find an interesting quandary that has become increasingly apparent in certain digital projects that cross my desk. Digital librarians have created and strive to adhere to certain benchmarks and guidelines for many aspects of the work, such as concrete digitization, and capture standards within national guidelines such as Federal Agencies Digital Guidelines Initiative (FADGI) and also from institutions like the National Archives (such as the Technical Guidelines for Digitizing Archival Materials for Electronic Access).

Other, more subtle aspects of the job are ultimately determined by the practitioner, or may be put to a working group or committee for final decisions. Some of the more perplexing areas, such as the more complex, murkier copyright issues or potential violations of privacy as result of digitization projects, have yet to be discussed widely. Perhaps these more difficult areas are harder to navigate and also imply a certain amount of risk, and we are arguably still in the early pioneer days of digital librarianship as a profession.

The aspect of privacy in digital collections is of particular interest, since the decisions practitioners make have very real and tangible impacts on the outside individual whose personal information may be at risk of disclosing, particularly in the era of broadcasting increasingly large amounts of digital media and information within connected networks with virtually perfect "memory."[1] The decisions around whether or not to digitize and publish content containing personal information in an online arena need to be weighed and examined as to potential risk and harm to the individual(s) impacted most directly by these actions.

Digital collections represent a digital version of analog collections and holdings (or are born digital content). These are materials that may be deemed rare, valuable, and/or of high potential interest for researchers. Candidates for digitization are accompanied with a number of factors that can impact or even impede dissemination—concerns of authorship, ownership, intellectual property, preservation, privacy, and so on. One rationale around digital collections is focused on digitization as preservation for aging materials, capturing a version of the original before any further loss occurs. In doing so, it prevents further handling of the original through its digital surrogate,

[1] There is a fascinating book entitled *Delete: The Virtue of Forgetting in the Digital Age* on the notion of memory and forgetting in a digital landscape by Victor Mayer-Schönberger (2009), who poses the idea of purposefully programming forgetfulness into technology, to provide a more forgiving, fallible (and human-like) network. And others like Pekka Henttonen call for more of a life cycle model, much like physical retention schedules, to set an expiration based on the orginal intent of information or a "purpose limitation principle" (Henttonen, 2017, p. 291).

although the wide dissemination presents its own set of challenges for the digital version. Copyright is one such complex area that could prevent distribution before the practitioner loads content into a wider arena of discovery. A copyright review will often take place before ingest and has been well documented and incorporated into production workflows, while privacy reviews have rarely been discussed widely for digital projects.

As digital librarians, our jobs often combine elements of project management and a certain level of technology skills and knowledge. This position often combines the more nuts and bolts aspects of the digitization and capture processes of digital media with more conceptual ones in understanding how digital platforms work, particularly in the indexing and ingesting functions. Depending on the institution, digital librarians can often feel that we work in a relatively lonely corner (depending on the size of your institution and the scope of its digital initiatives), in part due to the newness of the profession. These services often exist in a less traditional corner of library services, and some of us may be in the midst of reaching to full potential and ultimately could still be defining roles and services.

I have worked at two different academic libraries over the past decade, as the first person with the title of digital librarian at the institution, and previous to that, as a digital archivist within a museum setting. This aspect of defining one's role is, for the most part, a very freeing position, as it allows one to come into a brand new position and help mold expectations. However, a consequence of this situation has also meant that there has been a lack of direct mentorship that is often present in the more traditional library departments like Reference, Archives, or Cataloging (and in my case, getting the library science degree before digital library classes were offered). These are areas that have many decades of experience to impart on new hires as part of the training. As such, digital librarians face some interesting daily challenges that allow for a certain level of creativity and collaboration for the individual to craft the position and find answers to new problems as they arise, but lack the same level of standardization of practice that other areas of the library may already have in place. I also feel that the ethical areas can often lag behind the practice and the real world, particularly when the especially puzzling ethical conundrums present themselves, serving to be a very important piece of the long-term conversation and identifying themselves as areas in need of more discussion and awareness.

More recently, I have worked on a few specific digital projects that have had a higher potential of materials that contain private information. This is in part due to a grant which revolves around a modern archival collection. As such, many living individuals are represented in some way within the archival collections. The grant project has led to many frequent internal conversations to define exactly what constitutes private information within our institution, a process that had not been done before with regard to open digital collections. Digital librarians are, in one regard, the final gatekeeper of the information contained within digital collections. Practitioners decide when to upload and index collections to broadcast information to a wider arena than the former offline,

analog version allowed. "Exactly what information is developed and published is often the result of judgments and decisions made by numerous gatekeepers." (Mason, Mason, and Culnan, 1995, p. 264). As the final gatekeeper before wider dissemination, we are tasked with a responsibility to enforce both institutional policy (if available) and also be aware of more subtle, ethical issues that result as a consequence to our actions.

While the American Library Association (ALA) has addressed notions of patron privacy surrounding circulation records in the past, there has been a shortage of discussion and understanding revolving around the idea of privacy (and privacy violations) within digital collections at large. This has ultimately led to a disparity in practice within the profession, wherein decisions may often be made on the fly during initial project proposals and content selection regarding privacy, or perhaps not addressed at all. One of the more relatable units within the library that can be referred to for practice and policy around balancing issues of access levels and privacy is the archives. The archival community have long grappled with these issues in physical holdings, and how to best balance and determine access levels to more sensitive material (Gilliland and Wiener, 2014; Bingo, 2011; Benedict, 2003; MacNeil, 1992). In part, this is due to the fact that the originators of content (i.e., information producers) could be individuals, organizations, or families that may be unaware of the scope or existence of sensitive information within collections (Wood and Lee, 2015).

Privacy is a complex and often thorny issue present in modern society, with many varying opinions as to the scope and control of one's personal privacy and right (or expectation) to privacy. This issue is intriguing to me on a personal level since I have at times found myself in a moral conflict on this topic of privacy within digital collections, mainly between the work as gatekeeper to openly accessible digital collections, and as an individual who may have a higher level of expectation or desire of privacy in my personal life. I have been fortunate to have spent most of my professional work on projects that are, for the most part, freely and openly accessible digital collections with little to no risk of disclosing private information. The processes involved in translating analog collections into open, fully indexed digital collections is personally satifying to enable the dissimenation of information. Yet, at the end of the day, I leave my full-time job and have the option to live a *selectively* offline existence of relative anonymity in rural Northeast Ohio. I enjoy the peace and quiet that is attached to living in the remote countryside, affording at times a rather purposeful detachment from modern life. And in becoming a parent recently, I have come to realize that true sense of complete privacy, and the chance of a truly encompassing, offline existence for my children, is increasingly difficult in modern society (difficult, or perhaps impossible with the availability of a mountain of public data at the end of any Google search on any person). I share these personal aspects of my particular viewpoint here in the introduction, since I feel these ultimately play an important role in how I consider and make decisions to share information that is in the nature of the position of digital librarian. The decisions that practitioners make on a regular basis to publish and share widely content with private information will impact, and continue to impact, the lives of others as a direct

consequence of disclosing information online (and, as such, my role as the publisher of information becomes the *active agent* in the transformation from analog to digital containers).

Therefore, I have, within the role my position affords, a certain authority and capability to infringe on other's privacy by posting content to an openly accessible digital platform that may contain personal information. The process of digitizing analog collections and publishing online entails that items will be captured, described, and shared broadly to anyone with a solid Internet connection, in varying levels of discoverability through processes such as optical character recognition (OCR), transcription, and other methods to further liberate information from its physical container. As mentioned earlier, librarians have created selection criteria and related guidelines that serve to inform our work and justify decision making to build both physical and digital collections. Although as some of the real-world examples will show below, we oftentimes cannot fully comprehend the full implications of online dissemination and consumption, and as such need to be more mindful and aware of potential privacy disclosure.

Furthermore, as richly complex and complicated the notions surrounding privacy issues are, the related ethical decision making can be equally tricky and problematic. Ethical conundrums often fall into various shades of gray, with no easily definable right or wrong side to be found. Related decision making cannot often be put into fixed, predictive patterns due to these nuances and complexities. Each case can present new challenges and aspects to consider, and will therefore not likely fall into predefined, cookie cutter policies and/or frameworks. Frequently, such scenarios also fall outside of the range of legality, and can pose more of a moral dilemma for practitioners. Factors of permission or informed consent of content projected for online dissemination in digital collections can frequently be absent for consultation in project planning, and should be more at the forefront. And, finally, how can professionals become more keenly mindful of the impact and consequence(s) of digital publication of information? The first step is to create a framework or model for ethical decision making. Recently, some trends and conversations in conferences such as the National Forum on Ethics and Archiving the Web and projects such as student life documentation at the Barnard Archives and Special Collections[2] can point to more ethical considerations and treatment within the profession. The world of open access can at times conflict with individual privacy.

This title will strive to address concerns from both sides of the issue, acknowledging advocates of open access and privacy. I come into this debate as a practitioner who loves connecting the end user with resources by way of the free, open digital repositories, and as an individual who very much enjoys a private existence with a selective disclosure of personal information. The topic of privacy at large has been extremely intriguing as it often inflicts a dilemma at a fundamental level personally, with the desire to continue to build free and open digital collections while respecting

[2] Charlotte Kostelic presented on this topic at the 2017 Digital Library Federation (DLF) Forum about informed consent and care ethics on student life documentation project.

privacy boundaries of the individual. Ethical decision making can assist in this endeavor, finding solutions to best address the right to privacy balanced with the right to access.

As van Dijck (2013) noted in *The Culture of Connectivity: A Critical History of Social Media*, the boundaries between private and public life continue to be blurred with the rise of technological innovations and infrastructures. One such innovation is the digital library or repository, releasing the contents of an offline analog (or in some cases, born digital) into a larger realm of discoverability to a worldwide audience. Digital collections within these frameworks continue to build abundant amounts of information, providing the potential of connecting relatively disparate pockets of information on any given search query quite easily through the networked infrastructures.

> *Contemporary culture is fraught with its own debate over virtualism and the contradictions it presents. We live in a world where, as a result of digital communication technologies, we are increasingly connected to each other. These technologies distribute us, extending us in profound ways* (Miller, 2016, p. 2).

This aspect of a connective, limitless virtual world of information is wherein the potential privacy violation can exist.

As law professor Daniel Solove reflects, "The amount of personal information archived will only escalate as our lives are increasingly digitized into the electric world of cyberspace." (Solove, 2004, p. 26). The availability of data held throughout an endless number of networked and connected resources is staggering; content containing private information is oftentimes posted without an awareness of the individual in question. The enormity of the information readily available, wherein lives the potential for harm, increases as the virtual universe continues to grow exponentially. Introna (1997) has framed this convergence of public and private spaces via information technologies to challenge individuals by providing "fewer possibilities for making their existence their own." (p. 272). And Pekka Henttonen (2017, p. 287) ponders if in the end all spaces are potentially public ones, when examining all the ways individuals and others create digital information constantly.

The decision point for practitioners to release private information by way of the digitization and publication of the content is the central piece that directly impacts the overall scope and reach of content. One can also weigh the prospect of a nearly perfect memory inherent to digital networks of information (or Pekka Henttonen reflects on this as Bentham's panopticon in digital format, without spatial or temporal limits (2017, p. 290)), by way of the unrelenting technical structures with full index search capabilities. This idea can be directly juxtaposed with notions of programmable memory loss in computing (as suggested in the title *Delete*), or in the desire from the individual to remove content with their personally identifiable information as an act of erasure (as in The Right to be Forgotten case and subsequent ruling in Spain). A purposefully fallible machine designed to forget is just as fascinating and complex, but it is outside the scope of this discussion. Instead, the intent at this point is to make a call to digital librarians to begin to craft a framework

and model that can navigate varying levels of complexity related to privacy in digital collections, working toward a more holistic consideration of privacy in the role, and impact of, publishing digital content and information.

We will also explore some key concepts in the differing notions of privacy, and how ethical decision making can be applied as we continue to navigate these murky waters. Specific examples of how privacy may be present at different levels in digital projects will also be discussed. Recommendations will be provided as to how practitioners can better address these conundrums internally in a more conscious, concerted effort to address potential privacy breaches. Increasingly, complex questions have arisen in the ongoing publication of more and more content housed in open digital repositories. The enormous amount of information online can be overwhelming, and ultimately we practitioners need to hold ourselves accountable to make ethical decisions, and find a higher level of consistency in our role as stewards and gatekeepers of digital collections in this regard.