Articles

Of Provenance and Privacy: Using Contextual Integrity to Define Third-Party Privacy

Steven Bingo

Abstract

This article approaches the issue of third-party privacy by examining how contextual factors related to the creation and use of records can inform decisions to restrict or open access. Helen Nissenbaum's theory of contextual integrity, which originates from the discourse surrounding digital privacy, is applied as a means to expand an archival concept of provenance to address privacy risks. Applying contextual integrity to privacy decisions also allows archivists to frame decisions in terms of circulation, rather than as a simple dichotomy between access and restriction. Such nuance is invaluable when considering the impact of making records available digitally.

The central problem identified by many who have written about the protection of third-party privacy in manuscript collections is the lack of clear guidelines or principles that can be enacted on a profession-wide level. Because of the ethical dimensions of the privacy debate, concern for consistency from institution to institution is all the more pressing. At the center of this debate is what Mark Greene refers to as "the tension between access and property or privacy rights."[1] Bound up in this tension are concerns regarding the unintentional censorship of materials caused by restrictions on one hand and maintaining the trust of donors and third parties on the other.

[1] Mark A. Greene, "Moderation in Everything, Access in Nothing?: Opinions about Access Restrictions on Private Papers," *Archival Issues* 18, no. 1 (1993): 31–41.

Separate from the archival discussion is a discussion in the fields of computer science and information ethics regarding the privacy of digital information. One of the concepts to emerge from this discourse is Helen Nissenbuam's theory of "contextual integrity." One can begin to define information privacy rights, Nissenbuam argues, by understanding norms related to the context in which information is supplied, gathered, and used.[2] In other words, the norms of privacy surrounding a document may be determined by investigating a document's provenance. While this may seem obvious to archivists, emphasizing provenance as a tool to negotiate privacy concerns focuses the discussion toward appraisal, which has not been covered in much depth, other than to say that archivists should work with donors to identify and properly mediate risk.[3] Contextual integrity, as I will argue, provides archivists another tool with which to tackle privacy concerns in a more prospective, upstream manner. As suggested by some current literature, dealing with risk prospectively provides opportunities to make decisions at broad levels of organization.[4]

A second application of contextual integrity concerns access. Contextual integrity, Nissenbaum states, is violated when information divulged within one context is recast in another context, particularly of how the information is allowed to flow in radically different ways.[5] Nissenbaum cites the aggregation of consumer information gathered online as an example of how information provided in one context is appropriated in new contexts without the subject's knowledge.[6] As archivists embark upon mass digitization projects and seek out options for making born-digital documents publicly accessible, the question of reframing documents in new contexts becomes extremely pertinent.

After summarizing the current archival debate regarding third-party privacy, I will flesh out the specifics of contextual integrity. I will then articulate how contextual integrity translates into an archival concept of provenance and apply it to questions of appraisal and access. As a theory that incorporates both appraisal and access, I argue that contextual integrity can help align appraisal and access policies in a systematic and holistic fashion. I will also point out the limitations of Nissenbaum's theories within an archival setting that arise out of challenges unique to archivists regarding access and privacy. Specifically, contextual integrity does not resolve questions of privacy so much as it identifies key

[2] Helen Nissenbaum, "Protecting Privacy in an Information Age," *Law and Philosophy* 17 (1998): 559–96.

[3] Greene, "Moderation in Everything"; Frank Boles, *Selecting and Appraising Archives and Manuscripts* (Chicago: Society of American Archivists, 2005); OCLC Research, "Well-Intentioned Practice for Putting Digitized Collections of Unpublished Materials Online," rev. 28 May 2010, http://www.oclc.org/research/activities/rights/practice.pdf, accessed 01 July 2010.

[4] OCLC Research, "Well-Intentioned Practice."

[5] Helen Nissenbaum, "Privacy as Contextual Integrity," *Washington Law Review* 79 (2004): 119–58.

[6] Nissenbaum, "Protecting Privacy in the Information Age," 586.

factors that may bear upon the sensitivity of a document, such as the roles of the creator, recipient, and subject of a document. While I promote contextual integrity as a means of bringing privacy risks into focus, I also believe that the default position regarding restrictions should be on the side of access. In other words, it is important for the archivist to prove why a document presents a privacy risk great enough to override our duty as archivists to provide access. I present contextual integrity as a tool within a larger decision-making process informed by our professional ethics.

Literature Review

Archival concerns regarding privacy can be divided into legal concerns and ethical concerns. While the two are not mutually exclusive, the distinction is useful in understanding how the discourse surrounding privacy in archival collections is framed. Generally speaking, legal considerations are aimed at avoiding any punitive repercussions that might arise from the disclosure of sensitive or libelous information.[7] Ethical arguments, on the other hand, stem from a belief that a relationship of trust exists between the repository, the donor, and society.[8] While ethical arguments for the restriction of sensitive information often look to the law for general guidelines regarding privacy, they extend further to emphasize the importance of archivists as mediators of morally gray situations.

Most legal considerations regarding privacy revolve around broad definitions laid out by works such as Samuel D. Warren and Louis D. Brandeis's "The Right to Privacy" and William L. Prosser's "Privacy."[9] For Warren and Brandeis, privacy rights are an extension of "a right to life" that extends to the spiritual, emotional, and intellectual life of the individual.[10] Prosser further defines this right as consisting of four torts: intrusion upon one's solitude, disclosure of embarrassing private facts, publicity placing one in a "false light," and "appropriation, for the defendant's advantage, of one's name or likeness."[11] Specific

[7] Sarah Hodson, "In Secret Kept, In Silence Sealed: Privacy in the Papers of Authors and Celebrities," *American Archivist* 67 (Fall/Winter 2004): 194–211; Menzi L. Behrnd-Klodt, *Navigating Legal Issues in Archives* (Chicago: Society of American Archivists, 2008); Gary M. Peterson and Trudy Huskamp Peterson, *Archives and Manuscripts: Law* (Chicago: Society of American Archivists, 1985).

[8] Hodson, "In Secret Kept, In Silence Sealed"; Glen Dingwall, "Trusting Archivists: The Role of Archival Ethics Codes in Establishing Good Faith," *American Archivist* 67 (Spring/Summer 2004): 11–30.

[9] Examples of such works in archival discourse include *Navigating Legal Issues in Archives* by Behrnd-Klodt; *Archives and Manuscripts: Law* by Peterson and Peterson; *Privacy and Confidentiality Perspectives: Archivists and Archival Records,* ed. Behrnd-Klodt and Peter Wosh; and Hodson, "In Secret Kept, In Silence Sealed."

[10] Samuel D. Warren and Louis D. Brandeis, "The Right to Privacy," *Harvard Law Review* 4, no. 5 (1890): 193–220.

[11] William L. Prosser, "Privacy," *California Law Review* 48, no. 3 (1960): 389.

expansions to the protection of privacy with an impact on American archives are acts such as the Family Educational Records Protection Act (FERPA) and Title II of the Health Insurance Portability and Accountability Act (HIPAA).[12]

Many archivists observe that the general rights outlined by works such as those by Prosser and Warren and Brandeis are more likely to affect researchers than than they are archives.[13] For example, Greene argues that there is little evidence that restrictions based on broad legal definitions of privacy actually protect collections.[14] Instead, he proposes that archivists should only protect those classes of documents specifically outlined by legal statutes such as FERPA and HIPAA, as well as those identified in donor agreements.

Arguments like Greene's, which emphasize minimal intervention, serve at least two purposes. The first is to limit risk. Menzi L. Berhndt-Klodt observes that ambitious mandates to protect privacy place "a new affirmative legal duty and obligation [upon the archivist], and any failure to comply with such a duty or missteps in carrying it out may result in negligent conduct, possibly actionable."[15] The second purpose is to support broad access to collections. For example, Judith Schwartz argues that increased access to potentially taboo subjects can, if handled appropriately, play a positive role in advancing social justice. With respect to documents that expose queer and homosexual identities, Schwartz proposes that archivists and librarians "open the archives and research institutions to the full complexity of human lives."[16] In their work regarding the Mississippi State Sovereignty Commission, Sarah Rowe-Sims, Sandra Boyd, and H. T. Holmes illustrate how records that directly violate civil liberties might be responsibly opened for the public's benefit.[17]

While supporting access is a central part of the archival mission, access as an argument against privacy restrictions has its limits. Beyond the formalization of donor concerns within a donor agreement, some argue that archivists have a responsibility and interest in protecting privacy based on moral and ethical

[12] Menzi L. Behrnd-Klodt, in *Navigating Legal Issues in Archives*, provides a comprehensive list of statutes that impact access and privacy as of 2008.

[13] Examples include Hodson, "In Secret Kept, In Silence Sealed," and Peterson and Peterson, *Archives and Manuscripts* in the section regarding libel starting on page 44.

[14] Mark Greene, "The Power of Archives: Archivists' Values and Value in the Post-Modern Age" (presidential address, Annual Meeting of the Society of American Archivists, 2008), http://www.archivists.org/governance/presidential/GreeneAddressAug08.pdf, accessed 21 December 2009.

[15] Behrnd-Klodt, *Navigating Legal Issues in Archives*, 112.

[16] Judith Schwartz, "The Archivist's Balancing Act: Helping Researchers while Protecting Individual Privacy," *Journal of American History* 79 (June 1992): 179–89.

[17] Sarah Rowe-Sims, Sandra Boyd, and H. T. Holmes, "Balancing Privacy and Access: Opening the Mississippi State Sovereignty Records," in *Privacy and Confidentiality Perspectives: Archivists and Archival Records*, ed. Menzi L. Behrnd-Klodt and Peter Wosh (Chicago: Chicago: Society of American Archivists, 2005), 159–74.

grounds. Citing the Society of American Archivists' *Code of Ethics*, Hodson states, "The overall ethical tenet is clear: archivists must be aware of, and perhaps safeguard, the privacy of individuals represented in archival collections."[18] In his work about copyright in archives, Robert Cogswell observes that privacy is important not only for legal reasons (i.e., to avoid litigation), but also because "potential donors of useful records might withhold them if they suspect an archives might fail to protect their privacy."[19] Taking the ethical argument further, Timothy Ericson, in his case study regarding privacy in *Ethics and the Archival Profession*, argues that, "as a matter of conscience and morality," archivists should restrict private information that might cause undue harm to those close to individuals who might be compromised by materials in a collection.[20] If one believes, as Hodson argues, that privacy "more often constitutes an ethical concern than a legal one," then ethical guidelines regarding privacy need to address ramifications that extend beyond the courts and legal codes.

The hesitance accompanying the formation of ethical guidelines regarding privacy, I believe, stems from the ambiguity such a task is likely to involve. To restate Greene's concern, restrictions to access should not be made on the shaky premises of what one *might* find intrusive or embarrassing. The challenge, as Hodson and others describe, is to devise appropriate policies to assist collecting institutions in the navigation of an ethically gray area that will not require item-level examination. The scalability of appraising privacy risk is particularly pertinent in light of the increasing volume of documents resulting from digital technology and the relative ease of accessing and aggregating information on the open Web. For example, in 2010, an OCLC group dedicated to rights management created guidelines for the "well-intentioned practice" of placing archival materials online.[21] These guidelines include statements about evaluating risk at a collection or series level and attempting to obtain rights, when possible, at these broader levels of organization. With the numerous evolving challenges surrounding privacy and access that have been identified in the archival literature in mind, the following account of contextual integrity frames the theory as a general heuristic for evaluating potential risk couched within the context of the complexities posed by changing information technologies.

[18] Hodson, "In Secret Kept, In Silence Sealed."

[19] Robert Elzy Cogswell, *Copyright Law for Unpublished Manuscripts and Archival Collections* (Dobbs Ferry, N.Y.: Granville Publishers, 1992).

[20] Timothy Ericson, "Case Twenty-Nine," in *Ethics and the Archival Profession: Introduction and Case Studies*, ed. Karen Benedict (Chicago: Society of American Archivists, 2003), 63.

[21] OCLC Research, "Well-Intentioned Practice."

Contextual Integrity and Digital Information

> Conceptions of privacy that have served adequately until now are, in my view, unable to adapt to the new landscape, not quite able to conform to the ebb and flow of anxieties that these systems and practices provoke.[22]

This quotation by Nissenbaum reflects the impact that advances in media and information technologies—"the new landscape"—have on our concept of privacy and the concerns that the use of these advances raise. In the popular media, these concerns spawn stories over issues such as identity theft,[23] information sharing,[24] and unwanted exposure via social networking sites.[25] Within scholarly discourses, the debate regarding digital privacy extends to the disciplines of information studies, computer science, communications, business, and law.[26] These works range from articles describing technical solutions and problems,[27] to studies regarding privacy policies,[28] to ethnographic studies of how users of social networking sites mediate privacy concerns.[29] And, yes, even the concept of provenance has filtered into the field of data management as a means of determining authenticity and governing access to sensitive

[22] Nissenbaum, *Privacy in Context*, 148.

[23] Examples include T. Trent Gegax, "Stick 'em Up? Not Anymore. Now It's Crime by Keyboard," *Newsweek* 21 July 1997, 14; and Adam Cohen, David Jackson, Laura Locke, and Elaine Shannon, "Internet Insecurity," *Time,* 2 July 2001, 44+.

[24] Examples include Laura M. Holson, "Verizon Letter on Privacy Stirs Debate," *New York Times*, 16 October 2007, C1; and Deborah Branscum and Jennifer Tanaka, "Guarding Online Privacy," *Newsweek,* 5 June 2000, 77.

[25] Examples include Emily Gould, "Faith in Facebook," *Newsweek*, 21 July 2010, 9; and Dan Fletcher and Andrea Ford, "Friends without Borders," *Time,* 31 May 2010, 32–38.

[26] A cursory search of ISI Web of Knowledge performed on 24 May 2011 for the term "privacy and (online or digital or database)" in the topic field yielded at least thirty-two articles or papers within each of these fields, http://apps.isiknowledge.com/RAMore.do?product=WOS&search_mode=GeneralSearch&SID=1DGMJ1amP85egaKABpA&qid=1&ra_mode=more&ra_name=SubjectCategory&db_id=WOS&viewType=raMore.

[27] Examples include Umut Uludang, Sharath Pankanti, Salil Prabhakar, and Anil K. Jain, "Biometric Cryptosystems: Issues and Challenges," in *Proceedings of the IEEE* 92, no. 6: 948–60; Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan, "Private Information Retrieval," *Journal of the ACM* 45, no. 6: 965–82; and Saikat Guha, Kevin Tang, and Paul Francis, "NOYB: Privacy in Online Social Networks," in *Proceedings of the First Workshop on Social Networks* (New York: ACM 2008), 49–54.

[28] Examples include Margaret A. Winkler et al., "Guidelines for Medical and Health Information Sites on the Internet: Principles Governing AMA Websites," *Journal of the American Medical Association* 283, no. 12: 1600–1606; Anthony Miyazaki and Sandeep Krishnamurthy, "Internet Seals of Approval: Effects on Online Privacy Policies and Consumer Protections," *Journal of Consumer Affairs* 36, no. 1: 28–49; and Mary J. Culnan, "Protecting Privacy Online: Is Self-Regulation Working?," *Journal of Public Policy and Marketing* 19, no. 1: 20–26.

[29] Examples include Patricia G. Lange, "Publicly Private and Privately Public: Social Networking on YouTube," *Journal of Computer-Mediated Communication* 13, no. 1: 361–80; and Paul Dourish and Ken Anderson, "Collective Information Practice: Exploring Privacy and Security as Social and Cultural Phenomena," *Human-Computer Interaction* 21, no. 3: 319–42.

information.[30] From this interdisciplinary milieu, contextual integrity emerges as what Adam Barth et al. call, "a philosophical account of privacy in terms of the transfer of personal information."[31]

Within the archival realm, Heather MacNeil's 1992 book, *Without Consent: The Ethics of Disclosing Personal Information Held in Public Archives* represents an early attempt among archivists to proactively address threats to privacy posed by digital information.[32] More recently, both Sarah Hodson and Sarah Rowe-Sims et al. point out the troublesome dimensions of posting digital surrogates of personal diaries, correspondence, and photos to archives websites.[33] What is striking about the various conversations regarding privacy and digital information is the relevance of archival concepts, such as provenance, and the need for increased engagement between archivists and researchers in other disciplines attempting to tackle the challenges of privacy that have arisen in the past twenty-five years. Contextual integrity, because of its emphasis on the context, origin, and use of information, presents a promising overlap between archival discourse and literature regarding communication ethics.

In her 1997 article "Privacy in Public: Challenges in Information Technology," Nissenbaum uses "contextual integrity" to describe the social norms surrounding the use and dissemination of personal information.[34] Nissenbaum writes, "People count on . . . contextual integrity as an effective protection of privacy. Nightclub patrons may not mind being seen by other patrons but may reasonably object to having their actions reported outside of that context."[35] Here, Nissenbaum argues that privacy cannot be effectively defined within a dichotomy in which public information is "up for grabs" without any limits on access. Instead, privacy is defined by norms dependent upon the original context within which information is created and disseminated. While analyzing context to make access decisions is common practice among

[30] For the privacy angle, see Ragib Hasan, Radu Sion, and Marianne Winslett, "Introducing Secure Provenance: Problems and Challenges," in *Proceedings of the 2007 Workshop on Storage Security and Survivability* (New York: ACM, 2007), 13–18; and Qun Ni, Shouhuai Xu, Elisa Bertino, Ravi Sandhu, and Weili Han, "An Access Control Language for a General Provenance Model," *Secure Data Management* 5776: 68–88.

[31] Adam Barth, Anupam Datta, John C. Mitchell, and Helen Nissenbaum, "Privacy and Contextual Integrity: Framework and Applications," in *2006 IEEE Symposium on Security and Privacy*, 198–213, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1624011&isnumber=34091, accessed 9 July 2010.

[32] Heather MacNeil, *Without Consent: The Ethics of Disclosing Personal Information in Public Archives* (Metuchen, N.J.: Scarecrow Press, 1992).

[33] Sarah Hodson, "Archives on the Web: Unlocking Collections while Safeguarding Privacy," *First Monday* 11, no. 8 (August 2006), http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1389/1307, accessed 9 September 2011; and Sarah Rowe-Sims et al., "Balancing Privacy and Access," 159–74.

[34] Nissenbaum, "Privacy in Public," 207–19.

[35] Nissenbaum, "Privacy in Public," 215.

archivists, Nissenbaum's interrogation of a public/private dichotomy challenges the primary means by which many archivists adjudicate privacy concerns. For example, Elena Danielson describes archival methods for providing access as a "simple dual approach" where materials deemed public are open to all and materials deemed confidential are strictly closed.[36] In contrast, Nissenbaum argues that privacy is not simply a question of *which* information is permissible to share, but more a question of *how* information is shared.

Contextual integrity begins with a framing of context that, in many ways, reflects an archival concept of provenance. Specifically, Nissenbaum states that "Contexts are structured social settings characterized by canonical activities, roles, relationships, power structures, norms (or rules), and internal values (goals, ends, purposes)." Within this definition, one can see many parallels within archives, such as activities, roles (e.g., record creators), and power structures (e.g., organizational hierarchies). Even the concept of "internal values" finds correspondence with the notion of mandates that Terry Cook employs when describing macro-appraisal.[37] For example, providing for the health of a community may, in very simple terms, describe the mandate or goal of a hospital. While Cook argues that understanding such mandates should be central to appraisal, Nissenbaum states that these broad values and ends align the activities, roles, and organizational structures, which, in turn, inform specific acts of communication. Within any given context, one can begin to define privacy norms according to the type of information shared, the means by which information is shared (or should be shared), the people sharing the information, and those who are the subject of the information.

As an example, Nissenbaum cites the infamous recordings made by Linda Tripp, who taped a series of phone calls between herself and Monica Lewinsky. Among the pertinent contextual factors at play is the fact that the primary actors may be described as friends, one of whom happens to be the subject of the information in question. Furthermore, the type of information is sexual information of a personal nature that, if revealed, would likely be subject to broad exposure. Finally, Tripp recorded and disseminated the information to a third party (i.e., the investigators of the Clinton-Lewinsky scandal) without Lewinsky's knowledge. As Nissenbuam summarizes, "Even if, on balance, more good than harm came of Tripp's actions, they violated informational norms of friendship by transgressing transmission principles of knowledge, consent, and confidentiality."[38]

[36] Elena S. Danielson, "Ethics of Access," *American Archivist* 52(Winter 1989): 54.

[37] Terry Cook, "Mind over Matter: Towards a New Theory of Appraisal," in *The Canadian Archival Imagination: Essays in Honor of Hugh Taylor,* ed. Barbara Craig (Ottawa: Association of Canadian Archivists 1992), 38–70.

[38] Nissenbaum, *Privacy in Context*, 153.

While the intimate knowledge captured on Tripp's recordings may be defined by many as private in nature, it is also true that details of the affair were, at least initially, selectively shared with Tripp, if not with a circle of confidantes close to Lewinsky. This example illustrates a concept of social spheres, which works in correlation with contextual integrity. According to Jonathan Schonscheck, the idea of social spheres is based on the premise that each of us interacts among several social groups, each of which operates under varying social norms.[39] For example, Schonscheck points to one's relationship with one's spouse as different from one's relationship with a group of close friends, each bound by different expectations regarding behavior and communication. Drawing from Ferdinand Schoeman, Schonscheck argues that information disclosed appropriately in one sphere may seem inappropriate within another sphere.[40] For example, discussing one's sexual history with one's sexual partner may be appropriate, while discussing the same topic with one's employer is generally not appropriate.

Like Nissenbaum, Schonscheck's rendering of privacy is a response to the impact of digital communication technologies on privacy. Specifically, Schonscheck is concerned that the ability to aggregate information in the digital age will allow one to bring together information from many of an individual's social spheres. Schonscheck uses the following analogy:

> We can think of bits of information as tiles. The new compiling technologies are not packing tiles into boxes, nor are they strewing them in piles along the information superhighway. Rather, they are assembling them into mosaics.[41]

The concern here is that in an online context the public has the ability to aggregate through search engines several bits of information about any individual of interest. What's more, most individuals have little control over their personal information accessible online. An example by designer Paul Adams illustrates this point. He describes a girls' swim instructor's Facebook account that links her students with her circle of friends who occasionally participate in racy acts at their place of employment, a gay bar in Los Angeles.[42] Normally, the instructor would keep these spheres separate, but due to both the design of the site and the instructor's own ignorance, two realms of her life overlap. Depending upon one's sensitivities, this confluence of very adult behavior and regular interaction with children may raise concerns that otherwise would have been negotiated appropriately through the instructor's discretion.

[39] Jonathan Schonscheck, "Privacy and Discrete 'Social Spheres,'" *Ethics and Behavior* 7 (1997): 221–28.

[40] Schonscheck, "Privacy and Discrete 'Social Spheres,'" 223.

[41] Schonscheck, "Privacy and Discrete 'Social Spheres,'" 225.

[42] Paul Adams, "The Real Life Social Network v2," Slideshare, http://www.slideshare.net/padday/the-real-life-social-network-v2, accessed 23 August 2010.

In the example involving the swim instructor, as well as the Linda Tripp-Monica Lewinsky example, one could argue that contextual integrity was violated. In the case of the swim instructor, role-based norms governing her relationship with her students were violated by the intrusion of information from other realms of her life. Furthermore, her role as a friend may be compromised if she cannot control personal information in an acceptable manner. If a social networking site facilitates a violation of contextual integrity insofar as it allows for new flows of information not originally accounted for by responsible parties, then the question becomes one of assessing what kind of changes, if any, are warranted by the user, to the site, or to the site's policies.

As Nissenbaum points out, contextual integrity is useful for understanding information norms, and hence, for identifying instances when norms have been violated. However, Nissenbaum also states that contextual integrity on its own merits can be "conservative in possibly detrimental ways."[43] Specifically, contextual integrity does not interrogate how norms relating to information sharing can perpetuate unequal access to information. Nor does contextual integrity question whether the new context in which information is accessed may better support socially desirable ends. For example, the Freedom of Information Act represented an important change in the flow of information, a violation, as it were, of established norms, but one validated by its social benefits. As Nissenbaum argues, "Although contextual integrity refines our ability to identify when custom or expectation has been violated, and to predict potential sources of indignation, more is needed to assess the moral standing of custom in relation to novel practices."[44]

For Nissenbaum, the questions to ask beyond the maintenance of contextual integrity are twofold. The first is to consider the moral and political ramifications of altering the context within which information flows. This may include questions of fairness, power, and the general well-being of those affected. While this first consideration is extremely broad in scope, one can, with respect to archives, identify often-mentioned values that justify the movement of documents from private settings to public settings, such as supporting research and the expansion of public knowledge. The second consideration is how a change to an information flow "impinge[s] on the values, goals, and ends" of the context in which the information was intended. A current controversy to which this consideration has been applied is that of data mining in the field of law enforcement. Specifically, concerns have been raised over practices used to identify potential terrorists by identifying data patterns regarding travel and

[43] Nissenbaum, "Privacy as Contextual Integrity."

[44] Nissenbaum, *Privacy in Context*, 165.

telecommunications (where and to whom an individual makes phone calls).[45] While this is a clear violation of contextual integrity, proponents of the practice argue that potential gains toward the goal of preventing terrorist attacks outweigh this violation. The application of contextual integrity in this case depends, in part, upon how one evaluates the goals and value of data mining for the purposes of security. In this example, as in many other examples, contextual integrity does not define the ethically preferable path, but instead defines privacy risk, which, in turn, allows one to better weigh the benefits and consequences of changing the flow of information.

While Nissenbaum points out pertinent limitations to contextual integrity, she does not address one point particularly relevant to archives. Specifically, information shared within the context of a relationship marked by a high degree of trust and confidentiality is often very valuable for research because it provides insight into an individual's life that may not be available to the general public. The context of social spheres predicts this conflict between privacy and access to information, as the act of archiving inherently involves shifting documents from a relatively private context to a public context. For this reason, using contextual integrity as the sole rubric upon which to make access decisions regarding privacy is problematic and brings the archivist back to the conflict between privacy and access. As I have argued above and will point out in the next section, defining contextual factors related to privacy is a useful tool in identifying risk without having to delve into content, which, as pointed out in the appraisal literature, is a more time-consuming means of evaluating collections.

Contextual Integrity and the Appraisal of Privacy Risk

As described thus far, contextual integrity identifies privacy risks by framing norms regarding information flow within the context of the roles, activities, social structures, and goals shaping the creation of information. In applying these criteria concurrently with appraisal decisions, consider a collection of personal and faculty papers belonging to a writer at the institution where he taught. Typical appraisal criteria might identify value associated with the records creator's role as a faculty member and an active member in the local literary community, as well as the activities and relationships associated with those roles. Furthermore, the collection is likely to reflect, to borrow from Cook's theory of macro-appraisal, the overlapping of the "programme" of the university (i.e., its purpose and intent), the university as an agency or entity charged to carry out its program, and the individuals the university serves.

[45] Eric Lichtblau, "Study of Data Mining of Terrorists Is Urged," *New York Times*, 7 October 2008, http://www.nytimes.com/2008/10/08/washington/08data.html, accessed 14 September 2010.

While the value identified by appraisal criteria may justify the acquisition of the collection, these same criteria, when analyzed differently, can also identify risk. For example, one may begin by identifying roles, activities, and goals that pose little risk. The series that are largely associated with these factors would not bear a great deal of scrutiny in terms of evaluating privacy risks. On the other hand, roles, activities, or goals that are likely to pose privacy concerns should be identified, and applicable series should be subjected to a more rigorous analysis.

With regard to the writer, an archivist might, for example, consider that the writer's role within the writing community may place him in a position of confidence with respect to other writers in the community, a role that may be both personal and professional in nature. If the writer or his collection is of enough importance, then significant correspondence files may be identified either through conversation with the writer, the donor (if the writer is deceased), or friends within the department. By employing the concept of social spheres, one can evaluate with whom the writer has stronger ties and hence a greater likelihood of sharing sensitive information. Based on the importance of those with whom the writer is likely to have strong ties, one can evaluate the amount of time to dedicate to evaluating privacy risks and where that time might be most effectively spent. While a series or file of correspondence of high research value would merit the scrutiny to closely identify privacy risks, a series of marginal or unknown value might be lightly evaluated with an appraisal or access decision made based on a broad evaluation of risk against research value. In the latter case, the evaluation of risk would rest upon the functions or relationships that define the series. Series of questionable research value and a high likelihood of containing confidential information might be closed at the series level, as it would be difficult to justify the time needed to address immediate privacy concerns. Greene notes that while archivists should impose restrictions as a measure of last recourse, "the wisest course may be to close all or part of any collection which might *conceivably* represent an invasion of privacy."[46] Per the framework provided, I would replace "*conceivably*" with "likely based on contextual factors." As Aprille C. McKay points out, archivists need to define reasonable steps for identifying risks, including the levels of granularity at which risk may be appraised.[47]

In addition to helping archivists identify general activities or roles that might entail privacy risk, contextual integrity can also be a useful tool in working

[46] Greene, "Moderation in Everything," 37.

[47] Aprille C. McKay, "Third Party Privacy and Large Scale Digitization of Manuscript Collections: Legal and Ethical Obligations" (PowerPoint presentation from a panel discussion at "Extending the Reach of Southern Sources: Proceeding to Large-Scale Digitization of Manuscript Collections," Southern Historical Collection, University of North Carolina at Chapel Hill, 12 February 2009), http://www.lib.unc.edu/mss/archivalmassdigitization/download/mckay.pdf, accessed 20 July 2010.

with donors. As noted by Greene, Frank Boles, Behrnd-Klodt, and others, donor agreements provide a valuable tool for dealing with privacy risks. Although the benefits of working with donors have been well articulated, the literature does not provide much assistance regarding what questions one might want to ask. As Bill Landis points out, archivists can be far more proactive in developing best practices and tools when discussing issues of privacy and access with donors. Specifically, he notes that archivists should discuss both the positive and problematic implications of digitization, develop collective understandings or tools such as "access-restriction windows that might mitigate different categories of third-party privacy concerns," and develop "a more proactive approach to eliciting information about third-party privacy."[48] As I will argue, contextual integrity has the potential to help archivists most with respect to eliciting information about privacy.

Contextual integrity provides a number of avenues through which to broach the topic of privacy, specifically in terms of roles and activities of the records creator. For example, the archivist can ask directly if the records creator held any positions that provided access to private information or to individuals with whom the records creator had a close, informal relationship of trust. In addition to trying to identify relationships of trust and confidence that the records creator may have maintained, it might also be useful to ask how the records creator communicated sensitive information; in other words defining information flows. Such an activity could potentially begin with a rough mapping of the records creator's activities or relationships and lead to deeper probing of certain aspects of the collection if deemed worthwhile. Through an active donor/archivist collaboration, the donor's analysis of risk is sharpened by the archivist's understanding of provenance and functional analysis.

Because norms regarding privacy spring from contexts specific to the roles and spheres within which an individual interacts, discussions regarding privacy need not be isolated from a more general discussion regarding the records creator and the value of a collection. Should the discussion of a records creator reveal a position that might provide privileged access to personal information, perhaps as the treasurer of an organization or as an arbiter of family disputes, then this too may trigger discussions regarding potential third-party privacy concerns. By framing privacy as a quality that emerges out of specific social contexts, the archivist can deduce areas of potential risk arising out of general discussions about the records creator and discuss these risks with the donor. In this way,

[48] Bill Landis, "Reconciling Modern Archival Practices and Ethics with Large-Scale Digitization" (notes from panel discussion at "Extending the Reach of Southern Sources: Proceeding to Large-Scale Digitization of Manuscript Collections," Southern Historical Collection, University of North Carolina at Chapel Hill, 12 February 2009), http://www.lib.unc.edu/mss/archivalmassdigitization/download/landis.pdf, accessed 1 August 2010.

discussion of privacy need not be isolated from a general discussion of the records creator and the value of the collection.

Once risk has been identified according to roles and activities, which should translate to series-level risk, then decisions may be made regarding access, or what Nissenbaum would refer to as information flow. Applying contextual integrity to questions of access, I will refer to an example cited by Hodson—the Kinross papers. According to Hodson, Patrick Balfour, the third Baron of Kinross, was a travel writer whose correspondence contained "numerous letters pouring out intimate, confessional details."[49] Among these letters were the confessions of living gay men, some of whom were potentially closeted at the time the collection was processed.

In this example, contextual factors such as Kinross's role as confidant, the norms surrounding the activity of confession, and the goal of providing an intentionally limited outlet to the personal dilemmas of others point to the confidential nature of the correspondence. Furthermore, opening the correspondence changes the information flow surrounding the confidential information. These factors need to be weighed against the moral and political impact of providing access. To reiterate Schwartz's argument, I would point out that the correspondence may provide valuable insight into an underdocumented social history.[50] As stated above, factors such as the passage of time and the death of subjects named in the letters are also important. With the above arguments explicitly laid out, the archivist can make an appraisal of likely risk that supports either restriction or access.

In addition to addressing what materials should be made accessible, archivists need to consider how materials will be accessed. This question is particularly pertinent with regard to digitized and born-digital documents. In terms of digitized documents, the push toward large-scale digitization projects furthers the utility of large-scale appraisal decisions with respect to research value, intellectual property rights, and privacy risks. According to OCLC's document describing "well-intentioned practice,"[51] the selection of large-scale digitization projects involves weighing research value versus risk at a collection or series level. While, as stated earlier, series-level appraisal of risk need not only apply to digitization projects, the archivist has a wider range of potential options or consequences to consider.

In deciding access to a series like the Kinross correspondence, the archivist should consider at least three factors: when the materials will be open to the public, the conditions under which the materials will be made available, and how those materials will be presented to the public. The first factor is particularly

[49] Hodson, "In Secret Kept, In Silence Sealed," 200.

[50] Schwartz, "The Archivist's Balancing Act," 188–89.

[51] OCLC Research, "Well-Intentioned Practice."

pertinent if there are known privacy issues, as was the case with the Kinross papers. If the privacy risks surrounding a collection or series are not completely known, then the question becomes one of evaluating what types of access are reasonable given what is known after the collection is processed. This is where the second and third factors come more prominently into play. The second factor, which involves decisions related to conditions of access, may include tools such as user agreements. One example is the third-party privacy agreement in place at Duke University's Rare Book, Manuscript, and Special Collections Library.[52] While such instruments serve as a general protection against liability, they are also particularly useful for dealing with the uncertainty of opening collections that could not be processed at the item level. The final consideration involves decisions surrounding digitization and access to digital surrogates. Using contextual integrity, one could argue that a collection accessible on-site presents a lower degree of risk than collections freely accessible online. While access in a paper context allows a researcher to direct attention to sensitive information within a collection, access in an online context may require no such active intervention to make confidential information easily discoverable by all. In such a situation, controls like the agreement used by Duke University are less effective, for even if a researcher upholds the agreement, broad access to private information is still readily accessible. In addition, Nissenbaum and others argue that the problems of information flow in a digital context are not simply issues of scale, but also involve issues of aggregation. As such, concerns regarding the separation of social spheres become more pressing, particularly if access to content is made crawlable by large search engines through the use of optical character recognition software.

One other useful aspect of contextual integrity in terms of access is that it defines attributes, such as roles and activities, that may be used to support rule-based access.[53] Phoebe Evans Letocha, for example, argues that role-based access may be useful in determining appropriate access to medical records.[54] Based on rules that take into consideration laws like HIPAA, access may be automated so that electronic access observes legal and social norms. For example, an authorized medical professional could be provided full access to records within his or her home institution while different access privileges would be applied to researchers. This would be based on the roles attributed in a user

[52] Duke University Rare Books, Manuscript, and Special Collections Library, "Research Agreement," rev. 12 September 2011, http://library.duke.edu/specialcollections/services/dalton/research_agreement.pdf, accessed 12 September 2011.

[53] Barth et al., "Privacy and Contextual Integrity," 198–213.

[54] Phoebe Evans Letocha, "Contextual Integrity and Informed Consent: Providing Web Access to Images of Health and Medicine" (PowerPoint presentation, Annual Meeting of the Society of American Archivists, 2009), http://www.archivists.org/conference/austin2009/docs/Session502-Letocha.ppt, accessed 29 July 2010.

database and rules applied by programs governing access based on various factors including role.

Ultimately, contextual integrity provides an example of a structured means to evaluate privacy risk, one that can be applied to a number of different functions from appraisal to access. If, as Hodson argues, archivists bear an ethical responsibility to protect the privacy of third parties, then a set of standard criteria or tools can help assure both archivists and donors that a certain level of rigor is applied to privacy questions. Because privacy will always be subject to the archivist's judgment, our goal should not be the creation of hard and fast boundaries, but rather the identification of tools that will help evaluate risk and provide confidence that reasonable steps are taken to protect privacy. While contextual integrity is not likely to be a conclusive answer to archival concerns regarding privacy, it provides interesting possibilities for defining risk during appraisal that can be applied at broad levels of organization.