AMERICAN DOSSIER: YOUR LIFE ON THE INTERNET

Guy Philbin, Robert Morris University, philbin@rmu.edu
Debra J. Borkovich, Robert Morris University, borkovich@rmu.edu

ABSTRACT

This paper explores virtual privacy in the Information Age. It observes that personal data, transactional records, digital exhaust, and meta-data — provided with knowledge or without consent — are being gathered, compiled, stored, mined, and sold on the open market by governments, corporations, and individuals. These data are growing at an exponential rate in part due to naïveté, trust, and voluntary actions by technology users. Personal data are now routinely subjected to unprecedented intrusions as emerging technology has far outstripped any legal or constitutional protections. We conclude that relinquishing personal privacy is the currency spent in search of convenience and scarce time when using mobile computing devices, smart phones, and the internet. Our very human footprints, caches of personal and professional data, intellectual property, and private intimate details are manifested indelibly within our digital dossiers on the infinitely public internet.

Keywords: digital dossier, personal data, public data, privacy, security, data brokers, digital natives, digital immigrants

INTRODUCTION

According to Koehler [25], at the height of its power the East German secret police had amassed approximately six million dossiers on "enemies of the state" and "war mongering imperialists." Angwin [1], Baase [4], Bruce [10] and others have pointed out that from its formation in 1950 until it was dissolved in 1990, East Germany's state security service — the Stasi — was one of the most effective intelligence gathering secret police organizations the world has ever known. Yet even the Stasi could not have gathered the staggering amount of information on people's daily activities and personal habits voluntarily offered by millions of unwitting technology users and involuntarily collected by machines today. From the inception of the Information Age in the mid-20th century through the Digital Age of the 21st century, this paper illustrates how Americans willingly and unwittingly share their personal lives through voluminous bits and pieces of digital data — at times to their detriment and peril. Through the guise of a literature review we trace the construct of a digital dossier; and then follow-up our research with comparisons to several urban business journals' nonscientific surveys of individuals' awareness of sacrificing personal privacy and digital control. We conclude by examining the academic results of a 2013 Pew Research Center Report [34] that elicited adult responses regarding internet anonymity and privacy; and then reviewed a 2014 Pew Report [35] that canvassed subject matter experts' opinions and predictions on the future of the Internet in 2025.

THE FATHER OF DOSSIER AMERICA

J. Edgar Hoover deserves the dubious credit for being the modern father of Dossier America. As Weiner [54, p. xvi] observed in his history of the Federal Bureau of Investigation (FBI), "He was a founding father of American intelligence and the architect of the modern surveillance state." Nevertheless, the steady gathering of personal information by the government occurred well before Hoover and has continued unabated to this day. Citing Donohue [15] and Harris [21], Perrow [33, p. 87] has sketched the direction in modern times: "Though defunded by Congress, the Defense Department's Advanced Research Projects program with its eighteen data mining operations were transferred to the National Security Agency (NSA), CIA, and FBI." Perrow goes on to note that the assistance of all but one of the first-line telephone companies in NSA's data-mining programs became a political football in 2006 national electoral politics. In 2013 and 2014, the very same issues resurfaced with the revelations of former CIA employee and NSA contractor Edward Joseph Snowden. Hogan [22, p. 122] aptly framed this situation in a discussion of intelligence gathering nearly a decade ago when he observed:

> Data is manipulated daily to the detriment of the average person in America. Privacy is becoming a thing of the past. Virtually anyone in the world can get a credit report that lists all your debts, how quickly you pay them (or don't), and much more information related to your daily life.

In the 19th and 20th centuries, the power of the press — telegraph, radio, television, and analog computers — shaped public opinion through information and disinformation. By the late 20th and into the nascent 21st century, a new and more powerful tool arrived and thrived with Berners-Lee's 1989 development of the World Wide Web; and a new term was coined, "disintermediation — the elimination of middlemen" [55]. Dyson [17] proffered that: "The great virtue of the internet is that it erodes power… out of the center [taking] it to the periphery… giving to individuals the power to run their own lives." But by granting power to the people with the ability to create their own digital content and transmit it to whomever, wherever, whenever, we may have unwittingly sacrificed personal control, privacy, and safety by creating infinite online footprints of ourselves — our digital dossiers.

OPEN SOURCE HUMAN ENTITY RESOLUTION

The key to today's digital dossier is open source human entity resolution — a term used to describe the use of social media to monitor human networks. Pincus [36], for example, informs us that both the content and geographic location of tweets can be monitored in real time by an experimental computer tool named Raptor X. Developed by the Energy Department's Special Technologies Lab, Raptor was used by the Special Operations Command's National Capital Region (SOCOM NCR) to mine social media information as part of an experiment named Project Quantum Leap. As Pincus [36] pointed out, "What is stunning is that the project identified more than 300 traditional and nontraditional open sources as potentially relevant to the activity. These ranged from public sources such as the Patent Classification System, which has a lot of free business information, to subscription-based sources that sell specialized financial and business data." The fact that Raptor found over 300 open sources points to another disturbing fact, namely that there is currently no trustworthy way to distinguish this massively available private data from public information as both have grown well beyond the control of outdated privacy laws, simple rules, or opt-out mechanisms. Like Prism, another government surveillance program that collects video, e-mail, voice and stored data, technology like Raptor is redefining privacy in the modern world.

PUBLIC OR PRIVATE?

What began in 2006 as a small short message service (SMS) has become one of the most visited websites on the internet and a new metaphor for the blending of public and private information. Twitter has been used by popes and presidents, tourists and terrorists — and everyone in between — to send and receive all manner of public and private information. The blending of private and public information can be seen in news and entertainment stories actually created using information supplied by this technology. For example, CNN's Lopez and Gasts [26] based much of their report on a "string of tweets" rapper MC Hammer sent from his cell phone telling his side of a northern California traffic stop. Another example of how technology is shaping what we define as private information is the Supreme Court decision concerning cell phone searches handed down June 25, 2014. In a rare unanimous ruling, the Court recognized that cell phones are qualitatively and quantitatively different than other personal objects, that a search of digital information on a cell phone (including SMS data) is different than the physical search of a person, and that normally a search warrant should be obtained before law enforcement agencies search cell phones seized during an arrest [40].

Almost without realizing it, social media has become a primary source of personal and public information and the search and retrieval of socially enriched web archives using complex semantic queries is growing in availability and sophistication. Twitter content can easily be searched using engines such as Twitter Search, TweetScan, WeFollow, What The Trend, or Topsy. The information that can be derived from social media ranges from individual digital personas to news and politics that shape everyday life. Some view the millions of people posting to social media sites as a vast organic network ripe to be mined in real time. Sampling public opinion, predicting stock prices, tracking contagious diseases, marketing products, sourcing crowds offer a continually growing number of applications. As Spiliotopoulos, et al. [46, p. 174] have observed, "…from the moment that social networks such as Twitter provided an API [application programming interface] for collecting information, sentiment analysis can be performed in a multitude of ways." Smailovic, et al. [44, p. 77] have reported that "As more and more personal opinions are made available online, recent research indicates that analysis of online texts such as blogs, web pages and social networks can be useful for predicting different economic trends." Angwin [2, p. 91] has pointed out in a somewhat less rigorous way that sentiment analysis can be disquieting on a personal level: "Even creepier was a company called PYCO, which claimed it might be able to determine my personality type based on just my name and address. In its marketing materials, PYCO says it has created an 'algorithm to reverse engineer the data on a person's behavior — relationships, transactions, activities, interests, hobbies, purchase behavior, and so on.'"

PERSONAL IDENTITY INFORMATION

Anyone remotely interested in the glamour cinemas of the 1930s and 1940s, the FBI movies of the 1950s, and the James Bond films of the 1960s and beyond may recall the term "dossier" almost fondly. Evoking mystery, drama, the intrigue of WWII, the excitement of Cold War spies, and more modern special agents, dossiers represented information that often culminated in the dramatic resolution of good over evil. But in the real world, "dossier" evokes a more serious connotation. It is a file — paper, analog, or digital — of the most personal, private, and sensitive information available. For example, Pope Benedict XVI reportedly resigned after receiving one of the most memorable dossiers in recent times: a two-volume 300-page dossier, bound in red leather and embossed *Segreto Pontificio* [Pontifical Secret] that exposed corruption, blackmail, and homosexuality inside the Holy See. [47]

Unfortunately, there is hardly any area of our lives which remains private today. In spite of efforts like the Health Insurance Portability and Accountability Act (HIPAA) to delineate the privacy policies of health care services, personal medical information has never been more abundant. For example, personal data repositories like the MIB Group (formerly the Medical Information Bureau) contain extensive medical dossiers used for setting individual insurance rates and preventing fraud. Currently owned by roughly 500 insurance companies and in operation for more than 100 years, the MIB maintains millions of coded medical dossiers on Americans [57]. A statement linked to the MIB website claims: "It's not big brother or privacy invasion. Companies become members of MIB because it cuts their bad decisions and losses" [30]. Some of us might disagree with respect to privacy.

Personal privacy invasion is not the only area of concern within the scope of medical information. As with anything connected to a computer network or the internet, viruses and malware can infect patient monitors, lab analysis tools, surgical equipment and a host of other medical devices. According to Sun and Dennis [50], security concerns prompted the Food and Drug Administration to issue draft guidelines blocking approval of medical devices unless a manufacturer specifies how cyber security concerns are to be addressed. Considering the number of medical devices used for patient care that depend on electronics and software today, it was a decision made none too soon.

IDENTIFYING INFORMATION

In 2013 Target Corporation reported that it experienced a store credit card and bank card data breach that affected approximately 110 million customers. Although Target fired its CIO, expressed contrition, and displayed amazing transparency for a large corporation, the crisis nevertheless eroded customer trust and resulted in a very public revenue loss of 5.3 percent by quarter-end, as reported by *Barrons* on February 26, 2014. As news of Target's debacle reached critical mass in January of 2014, urban business journals in Atlanta, Dallas, Philadelphia, Portland, Puget Sound, and Tampa Bay [11] canvassed their readers as to whether or not security breaches like the one involving Target would prevent one from shopping with this retailer or a similar one. Only 56 percent of the 1,535 online subscribers of these combined journals who responded indicated they would be less likely to shop at a retailer with disclosed security breaches. This cavalier attitude on the part of the citizenry helps to explain why companies such as Target and Neiman Marcus Group Ltd. are successfully recovering from widely publicized data breaches.

While these results may reflect buyer reluctance to relinquish the convenience and expediency of purchase cards for a potentially more secure and anonymous method of payment, they are challenged by other studies which found consumers expect marketers to protect their digital identity and online privacy. One such study conducted in 2014 by New York-based Radius Global Market Research indicated that consumers feel the onus of data protection and personal privacy rests squarely on the shoulders of retailers and that buyers are willing to abandon brands that do not protect their personal information. "More than three-fourths of consumers indicated that they would stop doing business with companies that they felt had violated their privacy. A majority said that simply reading of hearing about a company's security breach makes them less inclined to buy/shop there (69 percent)" [39].

Perhaps even more troubling for customers concerned with personal privacy was a 2012 *New York Times* report detailing how retailers are eager to take financial advantage of very intimate personal information. Consumers going through major life events like the birth of a child, divorce, or changing jobs often don't notice or care that their shopping habits have changed. But as Charles Duhigg [16] made clear in his essay, for retailers like Target these life changes present a "holy grail" of marketing opportunities. It does not matter if this opportunity is exposed through grocery sales receipts, cell phones, e-mails, or a clever data mining algorithm. Life change events like Mom and Baby offer a sales bonanza for retailers — particularly if the seller can predict when a life change event is expected to occur — even when the consumer would rather not have anyone know.

INFORMATION SHARING

Government data mining is often done without consent. The information is not shared, rather it is taken on grounds of "national security." In this asymmetrical information exchange it can be argued that there is little or no personal benefit. The revelations of Mr. Snowden [29], who turned over thousands of classified documents to media organizations, created an international debate pitting national security against individual privacy. His revelations have drawn attention to a critical principle of big data: the more metadata, the greater its information value.

The U.S. Government and others are investing heavily in massive digital storage capacity, which has opened some amazing possibilities. Calls made and received, calls to others, calls associated with other phones, and the identity of even partial telephone numbers can be determined. The time, place, frequency, and duration of phone calls can be used to analyze traffic patterns. Computer networks, cell phones, and pagers add various other means of location to this detailed list of data. For example, if the channel access protocol (CAP) of a pager is known it is possible to record and view messages sent to that pager. In like manner, the media access control (MAC) used by most computer networks in conjunction with dynamic host configuration protocol (DHCP) lease data serves as a unique identifier to connect users [3]. Although the process of obtaining and capturing metadata is not standard because both manufacturers and service providers use different MAC configurations and deal with metadata in non-uniform ways, the amount of metadata available today has opened some fascinating new possibilities, as can be seen in the following request made by Rep. Steve Stockman (R-TX) to the House Committee on Government Reform and Oversight: "I respectfully request your Committee subpoena the records of every phone call made from all public and private telephones of all IRS personnel to all public and private telephones of all White House personnel" [48]. Alas, even if Stockman's request were acted upon, the requested metadata would not provide the desired information without sophisticated analysis.

INVOLUNTARY INFORMATION

We selectively share a great deal of information; it is particular information shared for a ad hoc purpose with some people but not others. Understandably, personal information is sometimes shared in return for a benefit: location for the ability to receive mobile calls; identifying information to use the Internet; or tracking data to produce ads targeting personal interests. There is, however, a growing concern with the information we are not sharing voluntarily. The NSA is reportedly collecting millions of facial images from social media, email, and text messages [41]. While facial recognition software portends a myriad of potential benefits, harvesting both public and private images will provide a trove of information years in advance of legislation or public debate that might consider personal privacy issues. This burgeoning harvest of facial-recognition information is not limited to intelligence agencies and the blurring of criminal and non-criminal databases has already occurred. As Timberg and Nakashima [52] have reported: "The faces of more than 120 million people are in searchable databases that state officials assembled to prevent driver-license fraud but increasingly are used by police to identify suspects, accomplices and even innocent bystanders in a wide range of criminal investigations." Among the FBI, State and Defense departments, it is conservatively estimated that there are currently more than 250 million facial photos. The lion's share of these images — 230 million — are from passports and visas.

HIDING INFORMATION IN PLAIN SIGHT

It is ancient wisdom that nothing in life is free. It is modern wisdom that "free" applications are often anything but free, as they can contribute substantially to the business of data brokering and to the automated gathering and exchange of public and private information. Consider an Android flashlight application that surreptitiously transmits geo-location data [54], Google bypassing Safari browser privacy settings to allow tracking [2], or any number of other disquieting breaches of virtual privacy that keep consumers in the dark while happily mining private information. In modern web mining usage, surreptitiously obtained data is analyzed and used to form a digital dossier. In cryptography, hiding information in "plain sight" is called steganography. Is there a real the difference between mining information using the code of an application without the user's knowledge and hiding information in an image? Both hide information, are surreptitious, intend to deceive and operate in plain sight.

One of the most impressive examples of technology that finds hidden things is made by Palantir, whose intelligence software has gained fame and fortune detecting fraud, spotting investments, increasing profits, rooting out terrorists for the C.I.A. and supporting U.S. military forces in combat. Founded in 2004 and named after a set of magic stones in *Lord of the Rings,* Palantir's headquarters is called The Shire. While The Shire is home to J.R.R. Tolkien's Hobbits, there is noting fictional about Palantir's clients which include J.P. Morgan Chase, Morgan Stanley, and state and local law enforcement agencies. Palantir is not alone in ferreting out information not meant to be known or seen by others. As Streitfeld and Hardy [49 p. B2] have observed: "New technologies like Google Glass are relentlessly pushing into territory that was out of reach until recently. From established behemoths to new startups, tech companies are bubbling with plans to collect the most intimate data and use it to sell things."

DATA BROKERS

Intimate personal data has become a commodity and there is considerable incentive for owners of websites or applications to covertly mine and sell or trade this information. Although owners and authors are legally responsible for what they release and there is a supposedly an obligation to make personal data anonymous there is nothing to prevent the trade or sale of this digital information. There are conservatively well over 250 personal data brokers ready to gather and sell basics like name, address, and phone numbers. As the number of firms entering this market continues to grow, other personal information such as age, net worth, hobbies, marital status, ethnicity, summaries of online social connections, and a host of other data including civil and criminal records, address history, family, relatives, and friends are becoming readily available. These so-called "people databases" include MyLife.com, Spokeo, US Search, PeopleFinder.com, and, of course, larger entities such as Yahoo, Google, Bing, Lexis-Nexis, Equifax, and Information America. If other firms using these techniques are aggregated, there are thousands of entities collecting, storing, and selling our personal information gathered through the internet.

The availability of personal information is growing apace with mobile devices which are proliferating at an astronomical rate. According to Cisco [14], mobile data traffic will increase 18-fold over the five-year period between 2011 and 2016. Online high-speed access and geo-position data have literally opened a picture window on people's personal preferences, social communications, habits, and daily activities. Moreover, as Kang [23, p. 1230] observed, private sector incursions into personal privacy are not protected by federal constitutional law. Governmental response to privacy concerns in the United States has been underwhelming to say the least. As Markoff [27] noted: "In agreeing to let private information brokers and credit reporting companies be governed by voluntary guidelines, the Federal Trade Commission is betting that companies that maintain vast computerized dossiers on people and businesses can police themselves in the face of ever more powerful technologies." Although some U.S. data brokers have agreed to third-party audits, privacy advocates are concerned that the protections offered by the Fourth Amendment will soon be meaningless. More recently, when Federal Trade Commissioner Julie Brill [9] was asked if personal data with names and personal identification was being harvested and used to make dossiers, her candid reply was: "Absolutely."

LOST INFORMATION

When information is prevented from producing other goods — from becoming information capital — it rapidly loses economic and social value. For example, GeoCities was the third most visited site on the web with roughly 38 million user-based pages when it was acquired by Yahoo for $3.57 billion in stock in 1999. Although the company had promised it would not release personal or demographic information to anyone without the users' permission, GeoCities apparently sold this information to advertisers. A consumer complaint against GeoCities resulted in a Federal Trade Commission consent order which prohibited the service provider from collecting and selling personally identifying information, as this was contrary to their stated privacy policy. (127 F.T.C., p. 94): "This consent order, among other things, prohibits GeoCities, a corporation that operates a World Wide Web site, from misrepresenting the purpose for which it collects or uses personal identifying information from or about consumers, including children." Shortly after this decision, Yahoo announced it would shut down the U.S. branch of GeoCities on October 26, 2009 and this popular web site joined a long list of Yahoo services in the internet service graveyard.

Economic fears that Europe's Data Protection Directive might preclude American companies from gathering data abroad are of greater concern to the domestic data brokering industry than are fears of invading individual privacy. Kang [25] tackled some of the issues raised here in "Information Privacy in Cyberspace Transactions," a paper prepared for the *Stanford Law Review*. In a nutshell, Kang attempted to set reasonable expectations and proposed a market solution for personal privacy that viewed information as a contractual commodity. Unfortunately, much as the Wired Equivalent Privacy (WEP) protocol is an outdated security algorithm [51], the velocity of change is such that the clear distinctions Kang drew between casual observation and surveillance have blurred in just a few years. Perhaps this is because of pure economics. According to Angwin [2], consumer tracking is the foundation of the $23 billion spent on online advertising in 2013. She backs up her claim that tracking is "exploding" by citing research conducted last fall by Worcester Polytechnic Institute and AT&T Labs which indicated that 80 percent of the internet's thousand most popular web sites are now using tracking technology — double the number in 2005.

ENDEARING TERMS OF PRIVACY

It is clear that Big Data and the Information Age have exacerbated the lag between technology and the rule of law with respect to privacy. A terms-of-service agreement — which is legally binding — is often subject to change. Privacy policies are often incorporated into service agreements. They are supposed to disclose how personal information is collected, stored, and released. Unfortunately, U.S. privacy laws apply only to the public sector and there is no general law when it comes to privacy. While some large companies have met the digital challenge by creating chief privacy officers (CPOs), conducting privacy audits, and adopting other measures to govern how they care for and share consumer information, others have not been nearly as transparent in their terms-of-service. As Baase [4, p. 105] frames it: "There, of course, continue to be many businesses without strong privacy policies and many that do not follow their privacy policies."

Do most consumers understand the internet as an advertising medium? Do most consumers know they are giving their consent to have dossiers assembled when they provide customer profile information? Or when they download an application to their smart phone that it will upload geophysical data on their whereabouts? Or that their smart phone tracks their whereabouts? Or are the economic forces and personal convenience driving online tracking and reporting of intrusive information just too attractive to resist? Currently, these arrangements are structured with service agreements often based upon the theory of half-life of information which originated with the Nobel-winning physicist Sir Ernest Rutherford's [42] work in radioactivity. Today this notion is expressed in the argument that some kinds of data — such as location data — are significantly less valuable as they age. Yet in terms of this analysis, the concept of a half-life no longer holds sway for two reasons. First, personal and public information is automatically being updated at a rate faster than most humans can comprehend. Second, as Mayer-Schonberger and Cukier [28] have observed, the reuse of data has shifted its economic value from its primary to its potential uses. Our research suggests that digital dossiers will exponentially increase in value as time passes; in most case negating the half-life theory as it applies to personal online information.

YOUR DIGITAL LIFE

Almost every major aspect of modern life is in some way touched by technological innovation and almost no human being — regardless of age or economic status — is entirely exempt. From the origins of the online bulletin boards of the 1970s, through widespread computer use in the 1980s, to the ubiquitous acceptance of the World Wide Web in the 1990s, the digital era has transformed our lives and the way we communicate with one another [32]. From the seminal underpinnings of government and academia, communities of users rapidly organized and for-profit corporations and non-profit organizations quickly glommed on. Commercialization of this enterprise boomed, bubbles were blown, and we were all strongly encouraged to join communities and share our personal and public information online.

Dependent upon interest and access to technology, terms like early and late adopters, technophiles and technophobes, digital settlers, techies, geeks and nerds appeared in common parlance; and shortly thereafter, in an attempt to stereotype their particular kind of reliance on interconnectivity and technology experience, entire generations of users were identified as "digital natives" or "digital immigrants." According to Prensky [38], the term "digital native" applies to the first generation of young adults born into the Information Age during the mid-1980s and beyond. They spend their lives surrounded by technology, using computers, videogames, mobile computing devices, smartphones and all the tools, toys and tricks of the Digital Age. Prensky [38] further asserted that those not born into the virtual world but have learned to adapt to the environment and to a certain extent adopt the technology are known as "digital immigrants." Without a doubt, "today's youth think and process information in a fundamentally different way from their predecessors" [38, p. 1]. Although they live and interact within a ubiquitous digital environment, digital immigrants still pride themselves in having some modicum of control over their own personal information; and to a much lesser degree, so do digital natives.

Other subject matter experts (SMEs) [6] vehemently disagreed with this tack, arguing that Prensky's metaphors generalized and marginalized both the roles of native and immigrant. Bayne and Ross [6] purported that Prensky lumped together all natives, regardless of gender, race, background, ethnicity, homeland, and cultural differences, presuming all were like-minded; and that all immigrants were akin to the biased 19th and 20th century stereotypes of heavily accented unintelligible foreigners. Selwyn [43] agreed that the terms digital natives and immigrants were too generic and did not speak to the true social-cultural aspects of individuals whose day-to-day lives, behavioral patterns, and habits also participated in elements of a very non-technological environment consisting of personal networks of family, friends, schools, community contacts, and reliance upon non-interactive print and other mass media sources for news.

NOT SO SAVVY

Selwyn [43, p. 364] found that young adults' use of digital technologies were often varied and unspectacular, asserting that the term digital native "highlights a misplaced technological and biological determinism." And O'Neil [31] argued to the contrary that not all young persons are universally digitally savvy; in fact, many proffer that turning a device on, making a call, swiping a finger on a notebook or smartphone, or starting an app may be all they need to know. Carr [13, p. 227] further argued that the effects of digital media cannot be described in generational terms and that "such distinctions strike me as misleading, if not specious… Net culture is not youth culture; it's mainstream culture."

So how do we describe children born into the 21st century? We are now entering a second generation of digital natives; perhaps they should be called, "Gen2 Digital Natives." They have no choice in technological participation; indeed Palfrey and Gasser [32] assert that children are digitized before birth when sonograms are taken in the womb and images are distributed from database-to-database, hospital-to-hospital, doctor-to-doctor, relative-to-relative, friend-to-friend, and even refrigerator-door-to-refrigerator-door. A figurative digital footprint of the unborn child emerges, already identified and tracked in a digital dossier constructed from data retrieved in utero. By the time a "Gen2 Digital Native" enters the workforce multitudinous digital files have been created and are available — from metadata to minutia — with thousands more to follow throughout the individual's lifetime.

Passports, licenses, credit cards, associations, memberships, bank accounts, and loans leave a trail. School, work, play, and travel identification is easily digitally secured. A steady stream of personal information is available to family, friends, colleagues, employers, marketers, advertisers, clubs, blogs, governments, and voyeurs of all ilk. Web cameras capture images, GPS sensors track locations, bar codes and RFID record what is bought or sold, VeriChips identify health issues. Physical characteristics are digitally harvested with assorted biotechnologies: facial recognition, fingerprints, eye scans, palm prints, walking and running gaits, DNA, x-ray, and countless other new and emerging technologies capture personal blueprints of each and everyone who is destined to live in the Information Age. Much of this technology, it turns out, is the basis of what the FBI calls next generation identification [18].

A "Gen2 Digital Native's" dossier thrives primarily upon his or her own digital contributions, but it also grows from interactive contributions of others' social media feedback. Constantly interconnected and networked the "Gen2 Native" prefers this method of communication and sharing information above all others. However, the convenience and excitement of being an integral part of a lightening-speed hard-to-control environment also means relinquishing personal control and privacy. Most people understand that technology acceptance requires trade-offs and willingly sacrifice control for convenience. But few realize the life-long impact of sacrificing significant privacy and the future implications of leaving eternal footprints in cyberspace. Palfrey and Gasser [32] liken these digital footprints as reminiscent of tattoos — something connected to a human being and difficult to be rid of as time passes. We wonder if confidence in the stewardship of corporate, organizational, and government databases in concert with unfettered participation in social media reflects a quiet trust or an immature naïveté?

TRUST VS. NAÏVETÉ

The concept of trust lies at the heart of every contributor to a digital dossier. A culture based upon openness, trust, respect, collaboration, shared meanings, and resources reflects a mutual understanding of goals, objectives and purpose. But how are these values and mission applied to "group-think" in a techno-virtual environment? In the 21st century, technology and culture do not lead separate lives, but are intimately connected implicitly and explicitly as they are linked by communities of practice, semiotics, signs, cues, and coded and decoded language that are digital, physical, sensory, oral, written, or analog in nature. The constructs of socio-technical theory [7] describes the interaction between people and technology in society, while social informatics [24] provides the tools of the interface.

Gleick [20] argues that most people (adults and youth, alike) believe that their online conversations, emails, blogs, e-commerce and banking transactions, and image postings are far more private and individually controlled than they really are. Access to information is now considered a birthright and our digital dossiers are virtually available for the buying or the selling, but more often than we care to admit, for free. All one needs to do is to access the internet, search, and then filter for results. Since electronic text is traditionally impermanent, revisions of digital information can be infinite thereby diminishing the pressure to achieve publication perfection [13]. The digital readers' reliance on and pleasure in informal and immediate access to information is based upon the perception that online data can be continually edited, updated, and enhanced. This attitude of digital dossier complacency encourages an uncritical trust of media sprinkled with the naïveté that database managers and textual editors will diligently and ethically take the time to correct items rather than moving on to the next headline.

Technology and culture tend to evoke dramatically different connotations, systems of meaning, experiences, and *Weltanschauungens,* the German expression for worldviews. Today's youth, college students, and many young adults purport that: "Technology is clean, powerful, exciting and a magical key to prosperity" [5, p. 1]. As naïve as this statement may be, some believe that technology can solve every problem. Technophiles [37] and early adopters [11; 12] argue that technology represents a necessary upheaval, innovation, and creative destruction to the permanence and stability of organizational, societal, and national culture. Our 21st century social-cultural environment coupled with disruptive technological innovations perpetually inspires our will to stay connected and our willingness to contribute to our ever-expanding digital dossiers. The need to be known, to be out there, to be meaningful, to matter in the digital reality, may be the "Gen2 Digital Native's" true raison d'être.

IDENTITY AND TIME

Breese-Vitelli and Borkovich [8] asserted that digital natives are more likely to use robust and flexible information systems including mobile applications and various social media such as WIKIs, YouTube, Facebook, Twitter, and others for sources of shopping, news, directions, and related information often without checks for reliability, credibility, and accuracy. As Gleick [20, p. 410] so eloquently put it: "When information is cheap, attention becomes expensive." Digital natives are clearly more willing to accept the risks associated with lack of data confirmation and the loss of privacy and control when participating in online social media, applying for credit cards, jobs, loans, e-shopping, e-dating, posting and sending personal emails and photos, and many other tasks. Turkle [53] described this phenomenon as a perfect storm with users obliviously working in the still center. She explained that we are overwhelmed by the thrill of so much available information and our contributions to it that we are drawn to digital connections that appear to be low risk — but in reality these relationships expose our vulnerabilities. Turkle [53] further posits that we expect more from technology and less from each other, falsely presuming that technology will take care of us. At times, users are completely oblivious that they are continually contributing to their own digital footprints — their digital dossiers.

In *Born Digital*, authors Palfrey and Gasser [32] argue there is a distinct difference between one's identity and one's digital dossier and that it behooves all technophiles and technophobes to learn and understand these salient and material distinctions. The authors describe one's identity in personal and social characteristics, assert that one may wish to disappear within a crowd; or alternatively, have multiple identities, online and offline. "Young people disclose information about themselves online to build trust with others and [to extend] their lives offline" [32, p. 25]. The authors liken this youthful online obsession to their parents' generational pursuits of endless talking on the phone and hanging out at malls or fast-food joints. Palfrey & Gasser [32, p. 39] advocate digital literacy skills, education, and common sense to control online content so when an individual's "personally identifiable information or PII" is compiled into a digital dossier it can be managed and protected. For Solove [45], the sum total of collected digital information held from many different sources at any time, about any person, makes up his or her digital dossier. Once compiled, this dossier continually expands with new data; and as Angwin [1] has demonstrated it is difficult, if not impossible, to modify, revise, change, or delete information. These subject matter experts advise online users not to confuse establishing their personal and social identities with creating permanent online dossiers.

The persons who can do the most to protect their digital dossier privacy over the long haul are the Digital Immigrants, Natives, and Gen2 Natives. Common sense over disclosure of potentially sensitive, harmful, damaging, or private information about oneself or others remains the primary responsibility of the discloser. Social media and other websites continually publish disclaimers, routinely alter their access and security policies, typically share data with marketers and advertisers, and update their privacy rules at will. Until legislation catches up with technology — if ever — the prudent man or woman rule for the actor, the user, still applies. Even with constant vigilance, our digital dossiers are available for distortion, manipulation, amusement, and yes, on occasion, they are even capable of providing credible information for legitimate searches. Unfortunately, much of the data in digital dossiers that were once personal intellectual property dependent upon a database and its security are now considered public information residing in the virtual public domain.

KEY ISSUES

Our research pointed to several key issues. Most important is that the majority of people haven't a clue about how the technology works or what happens to the data they use every day. However, interesting anomalies existed when our research was compared to the results of nonscientific business polls and surveys. Summary results of these urban business polls and surveys are available in the Appendix.

For example, in 2014 the *Pittsburgh Business Times (PBT)* posted an online survey asking readers if they were even attempting to protect their personal information on the web [11]. Some 68 percent believed they were not only aware of internet dangers, but had taken positive steps to protect their data. In 2013, the *Atlanta Business Chronicle (ABC)* asked online readers if they trusted entities and individuals who have access to their private information [11]. With a no-confidence vote of 91 percent, *ABC* subscribers overwhelmingly responded that they perceived government agencies and corporations to be seriously untrustworthy. The majority of *PBT* and *ABC* subscribers are adult college graduates, academics, entrepreneurs, large business corporate sponsors, and non-profit directors and managers. These demographics unfortunately do not differentiate among gender, age, discipline, or corporate position. The *PBT* portrays its respondents as primarily in control of their data; the *ABC* poll clearly reflects a lack of trust in government entities or corporations that maintain massive databases. However, these data do not reflect our research indicating that students and young adults continue to exhibit a high degree of trust and naïveté when it comes to online participation. As these polls do not indicate what type of protective measures the respondents have implemented, follow-up research using a wider audience to verify their respective results, is warranted.

From an academic research perspective, the 2013 PEW Internet & American Life Survey titled, "Anonymity, Privacy, and Security Online" [34], sampled 792 adult internet and smartphone users with the same types of queries. According to PEW, a clear majority (86 percent) of users have taken steps to remove or mask their digital footprints ranging from clearing cookies (64 percent), using fake names (18 percent); to encrypting email (14 percent). It is also clear that the frequency of internet access by "digital natives" vastly outpaces that of "digital immigrants" and that American adults recognize that they are being victimized by online identify theft and malicious or mischievous privacy invasions [34]. Although such nonscientific business polls often conflict with current academic research, it is important to note that we are — and may continue to be — in dire peril regarding the control and privacy of our personal online data. More to the point, according to Pew, our digital dossiers are likely to experience tumultuous change over the next decade.

LOOKING FORWARD

The Pew Research Center Report (2014), "Digital Life in 2025" [35], marking the 25th anniversary of Berners-Lee's creation of the World Wide Web, canvassed over 2,500 privacy, cyber-security, and net neutrality SMEs about the future of the web. Some of their predictions included the state of digital life in the year 2025 — and not surprisingly several theorized about the personal dangers of what we term the digital dossier. One anonymous contributor suggested that: "People will continue — sometimes grudgingly — to make tradeoffs favoring convenience and perceived immediate gains over privacy; and privacy will be something only the upscale will enjoy" [35, p. 1]. Regrettably, a significant number of survey respondents said most of the internet public will mindlessly exchange their personal information or future freedom in some regard for something they find attractive in their near-term interests. However, the overall expert consensus agreed that "the Internet will become 'like electricity' over the next decade — less visible, yet more deeply embedded in people's lives, with many good and potentially bad results" [35, p. 1]. Even more disturbing, Pew, in concert with Elon University [35], further reported that key themes emerged from 1,500 SME respondents predicting consequences for the Internet in 2025. The following theories augur that we may encounter more problems than solutions:

> Dangerous divides between the haves/have-nots resulting in resentment and possible violence; Abuses and abusers will evolve and scale; Loss of privacy; Persons may be tracked/watched/recorded without knowing it; Governments will assert power invoking security/cultural norms; Humans/Organizations may not respond quickly to challenges imposed by complex networks; Communication networks will be more disruptive; People will be connected all the time in the sense that no one remembers what it was like to be disconnected; People will lack critical thinking, information literacy skills; Illnesses will surface based on anxiety, stress and being connected all the time; *People will be unable to manage their digital identities*. [35, pp. 9-12, Abridged]

From the discovery of fire and the wheel, through "farms, factories, and floppies" [16, p. 14; 48] every technology has resulted in both positive and negative cultural, social, political and economic impacts. The evolution of the Information Age is no different. We have changed, adapted, overcome, and thrived. Is there any reason to believe we won't survive this latest personal internet challenge?

CONCLUSIONS

Like the internet, our digital dossiers are infinite caches of information. As Max Frankel [19, p. 5] has pointed out, "Information that is gathered and managed in secret is a potent weapon — and the temptation to use it in political combat or the pursuit of crimes far removed from terrorism can be irresistible." The control, safeguarding, and protection of our personal data — our intellectual property — requires further consideration, study, and reasonable lawful solutions. It seems likely that passwords will be replaced by biometric identifiers (like the iPhone home button that recognizes a thumb print) and these will become the new keys to guard our most personal private information. It also seems likely a day will come when a positive identification could be made in the time it takes to walk by a video camera. Notwithstanding, our digital identity is not merely a virtual "second" life in cyberspace. A digital dossier is much more than an online resume of personal, public, and transactional data. *It is your private life — on the infinitely public internet.*

REFERENCES

1. Angwin, J. (2014). *Dragnet nation: A quest for privacy, security, and freedom in a world of relentless surveillance*. New York: Times Books.
2. Angwin, J. (2014, July 30). The web's new gold mine: Your secrets. *The Wall Street Journal.*Retrieved from: http://online.wsj.com/news/articles/SB10001424052748703940904575395073512989404
3. Balakrishnan, H. (2005.) Wireless channel access protocols. Retrieved from: http://nms.csail.mit.edu/6.829-f05/lectures/L11-wlessmac.pdf
4. Basse, S. (2008). *A gift of fire*. New Jersey: Upper Saddle River, Pearson Prentice Hall.

5. Batteau, A. W. (2010). *Technology and culture*. Long Grove, IL: Waveland Press, Inc.
6. Bayne, S., & Ross, J. (2007). The 'digital native' and 'digital immigrant': A dangerous opposition. *Annual Conference of the Society for Research into Higher Education*(Paper Presented), 1-6. Retrieved from: http://www.malts.ed.ac.uk/staff/sian/natives_final.pdf
7. Bostrom, R. P., & Heinen, J. S. (1977). MIS Problems and failures: A sociotechnical perspective - Part I: The cause. *MIS Quarterly, 1*(3), 1977, 17-32.
8. Breese-Vitelli, J., & Borkovich, D. J. (2013). Mobile technology culture and its impact on college students' local news viewing behavior. *Issues in Information Systems, 14*(2), 400-410.
9. Brill, J. (2014, March 9). The data brokers: Selling your personal information. *60 Minutes Report by Steve Croft*. Graham Messick and Maria Gavrilovic [Producers].Retrieved from: http://www.cbsnews.com/news/the-data-brokers-selling-your-personal-information/
10. Bruce, G. (2003). The prelude to nationwide surveillance in East Germany: Stasi operations and threat perceptions, 1945–1953. *Journal of Cold War Studies, (5)*2, 3-31.
11. *Business Polls and Surveys*. Retrieved from: www.bizjournals.com/poll/archives
13. Carr, N. G. (2010). *The shallows: What the internet is doing to our brains*. New York: W. W. Norton & Co., Inc.
14. Cisco. (2012). Cisco visual networking index forecast projects 18-fold growth in global mobile internet data traffic from 2011 to 2016. Retrieved from: http://newsroom.cisco.com/press-release-content?articleId=668380
15. Donohue, L. (2006, January 12). You're being watched. . . . *Los Angeles Times*. Retrieved from: http://pqasb.pqarchiver.com/latimes/doc/422164121.html
16. Duhigg, C. (2012, February 16). How companies learn your secrets. *The New York Times*. Retrieved from: http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all&_r=0
17. Dyson, E. (2008, November 18). It sucks power out of the center: Does Google violate its 'don't be evil' motto? *Intelligence Squared Debate*. Retrieved from: www.npr.org/templates/story/story.php?storyId=97216369
18. Federal Bureau of Investigation. (2014). Next generation identification. Retrieved from: http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/ngi
19. Frankel, M. (2013, June 23). "Where did our inalienable rights go?" *The New York Times*. Sunday Review, p. 5
20. Gleick, J. (2011). *The information: A history, a theory, a flood*. New York: Vintage Books.
21. Harris, S. (2206, February 23). TIA lives on. *National Journal*. Retrieved from: http://shaneharris.com/magazinestories/tia-lives-on/
22. Hogan, K. (1996). *The psychology of persuasion*. New York: Pelican Publishing Company.
23. Kang, J. (1998). Information privacy in cyberspace transactions.*Stanford Law Review, 50*(4), 1193-1294.
24. Kling, R. (1999). What is social informatics and why does it matter? *D-Lib Magazine, 5*(1). Retrieved from: http://www.dlib.org/dlib/january99/kling/01notes.html.
25. Koehler, J. (2000). *Stasi: The untold story of the East German secret police*. New York: Westview Press.
26. Lopez, E., & Gast, P. (2013, February 24) "MC Hammer arrested in obstructing officer case." Retrieved from: http://www.cnn.com/2013/02/23/showbiz/california-mc-hammer-arrest
27. Markoff, J. (1997, December 18). Guidelines don't end debate on internet privacy. *The New York Times*. Retrieved from: http://www.nytimes.com/1997/12/18/us/guidelines-don-t-end-debate-on-internet-privacy.html
28. Mayer-Schonberger, V. &Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.
29. Mazzetti, M & Schmidt, M. (2013, June 10). Ex-C.I.A. worker says he disclosed U.S. surveillance: Revealed data trove. *The New York Times*, pp. A1, A13.
30. MIB Group, Inc. "Is the MIB the life insurance bad guy or good guy?" Retrieved from: http://www.hinermangroup.com/blog/life-insurance-faq/mib/
31. O'Neill, M. (2014). Confronting the myth of digital native.*Chronicle of Higher Education* (21 April 2014). Retrieved from: http://chronicle.com/article/Confronting-the-Myth-of-the/145949/
32. Palfrey, J., & Gasser, U. (2008). *Born digital: Understanding the first generation of digital natives*. New York: Basic Books.
33. Perrow, C. (2007). *The next catastrophe*. Princeton, New Jersey: Princeton University Press.
34. Pew Research Center. (September, 2013). *"Report: Anonymity, Privacy, and Security Online."* Retrieved from: http://www.pewinternet.org/Reports/2013/Anonymity-online.aspx/
35. Pew Research Center. (March, 2014). "*Report: Digital Life in 2025."* Retrieved from: http://www.pewinternet.org/2014/03/11/digital-life-in-2025/
36. Pincus, W. (2013, August 7). Mining social media: The new way of life. *Washington Post*. Retrieved from:

http://www.washingtonpost.com/world/national-security/mining-social-media-the-new-way-of-life/2013/08/07/a6878fc6-fed4-11e2-9a3e-916de805f65d_story.html
37. Postman, N. (1993). *Technopoly: The surrender of culture to technology*. New York: Vintage Books.
38. Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon, 9*(5), 1-6. Retrieved from: http://www.marcprensky.com/writing/prensky
39. *Quirk's Marketing Research Review.* (July 2014). "Data privacy: Onus is on us. Consumers unsatisfied with current state of information protection." *XXVIII*(6), 18.
40. Riley v. California, 13-132 (2014).
41. Risen, J & Poitras, L. (2014, June1). N.S.A. collecting millions of faces from web images. *The New York Times*. pp. A1, A19.
42. Rutherford, E. (1908). Volume of emanation. *Philosophical Magazine, 16*(92), 300-312.
43. Selwyn, N. (2009). The digital native – myth and reality. *Aslib Proceedings: New Information Perspectives, 61*(4), 364-379.
44. Smailovic, J., Grcar, M., Lavrac, N., & Znidarsic, M. (2013). Predictive Sentiment analysis of tweets: A stock market application. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data - Third International Workshop Proceedings Series*, South Maribor, Slovenia: Springer, 77-88.
45. Solove, D. (2004). *The digital person.* New York: New York University Press.
46. Spiliotopoulos, D., Tzoannos, E., Cabulea, C., & Frey, D. (2013). Digital archives: Semantic search and retrieval. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data - Third International Workshop Proceedings Series*, South Maribor, Slovenia: Springer, 173-182.
47. Stanglin, D. (2013, February 22) Report: Pope resigned in wake of gay priest scandal. *USA Today.* Retrieved from: http://www.usatoday.com/story/news/world/2013/02/22/pope-leaks-fallout/1938321/
48. Stockman, S. (2013, June 11). "Stockman requests subpoena of NSA's White House, IRS phone logs." Press release. Retrieved from: http://stockman.house.gov/media-center/press-releases/stockman-requests-subpoena-of-nsa-s-white-house-irs-phone-logs
49. Streitfeld, D. & Hardy, Q. (2013, June 10). "Data-driven tech industry is shaken by online privacy fears." *The Wall Street Journal*, B1, B2.
50. Sun, L. & Dennis, B. (2013, June 14). "FDA moves to protect medical equipment." *The Washington Post*, pp. A5, A16.
51. Tews, E. (2007). Attacks on the WEP protocol. (Unpublished thesis). Department of theoretical computer science. Technical University Darmstadt, Hessen, Federal Republic of Germany.
52. Timberg, C. & Nakashima, E. (2013, June 17). "Photo-ID databases become troves for police." *The Washington Post*, A1, A5.
53. Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books.
54. Vincent, J. (2013, December 6). Free Android app stole location data to send to advertisers. *The Independent*. Retrieved from: http://www.independent.co.uk/life-style/gadgets-and-tech/free-android-flashlight-app-stole-location-data-to-send-to-advertisers-8988668.html
55. Weiner, T. (2012). *Enemies: A history of the FBI.* New York: Random House.
56. Winer, D. (2005, December 7). Dave Winer's blog. *Scripting News*. Retrieved from: http://scripting.com/2005/12/07.html
57. Zall, M. (1996, April). Medical records bureau agrees to open files. *Nation's Business, 84*(4).

APPENDIX

Table 1. Pittsburgh Business Times Poll
Business Pulse – Online Poll (03-28-14)

Q: How serious are you when it comes to protecting your personal data?	
Not at all but I should be more careful	20%
There's nothing I can do about it	12%
I get identity theft protection, use others	68%
Votes Cast = 127	100%
This survey is not a scientific sampling, but offers a quick view of what readers are thinking.	

Table 2. Atlanta Business Chronicle Poll
Business Pulse – Online Poll (09-22-13)

Q: Whom do you trust most with your private information?	
National Security Agency (NSA)	4%
Internal Revenue Service (IRS)	5%
FACEBOOK	0%
I don’t trust any of them.	91%
Votes Cast = 475	100%
This survey is not a scientific sampling, but offers a quick view of what readers are thinking.	

Table 3. PEW Internet & American Life Survey
“Anonymity, Privacy, & Security Online” (09-05-13)

Q: Have you experienced internet issues that compromised your online privacy? A: Yes, based upon the following age groups:			
18 – 29	30 – 49	50 – 64	65 +
55%	42/%	30%	24%
Q: How was your online privacy compromised?			
Email or Social Media Hacked			21%
Stalked or Harassed Online			12%
Important Online Information Stolen			11%
Victim of Online Scam & Lost Money			6%
Encountered Physical Danger			4%
Reputation Damaged Due to Post			6%
Sample Size (n) = 792 Adults Aged 18 or Older Margin of Error = + / – 3.8%			n=792