

## **The Data Mine: *Whose Gold is it?***

DAVID PIKE

### **Introduction**

Research for this paper began with the idea that personal information is increasingly, if not entirely in digital format and is being collected and potentially indefinitely retained by entities outside the *self*.

Other authors such as Viktor Mayer-Schönberger<sup>1</sup> and Jean-François Blanchette and Deborah G. Johnson<sup>2</sup> have effectively illustrated both the personal and social value of degrees of forgetfulness across one's life. In summary, they convincingly frame the aspects of "forgetting" being the historical human default as opposed to our contemporary condition that, via digital information replacing human minds and paper records, remembering and gross data retention are now the default. They contrast the panoptic effects of potentially everything being known about you, against the Western cultural tradition of being capable of having a personal "chance for a new start." They also encourage and offer

methods and for certain aspects of retained personal data to be “forgotten” and/or never retained in the first place.

What I would like to take up is the value of this information. I propose that there is a dilemma based around not only the value of personal information to the agents involved, but also the ownership of this information. I focus on personal data captured and archived via private sector or commercial activities, and begin by framing these capture events in the context of an *exchange*. This Exchange Concept is based on the idea that information is acquired either directly or indirectly via an exchange between parties. Sometimes the exchange is very equitable, sometimes far from it. Ethically, I suggest that there are Justice and Rights issues. Justice extends from the notion of John Rawls’ concepts of “justice as fairness”, in general that “basic liberties have priority over aggregate social welfare...that fair...opportunities be provided equally for all citizens...”<sup>3</sup> Rights in this case, are ownership or property rights. As a modern extension of Locke’s concept that one has a claim to property when they “mix their labor’ with a thing.”<sup>4</sup> This ethical notion of Contract Theory is used to propose evolution of existing “contracts.”

From this I offer a perspective regarding the valuation of personal data to both individual and the entity either capturing or using the retained data: How is it that the stakeholders find value in the information? First, I argue that to the capturer/user, this information has become too valuable and entrenched in cutting-edge business models to be as readily “forgotten” as the previous authors suggest. Second, there develops an inequity after the initial exchange, as data accumulates and the individual’s “picture” becomes increasingly valuable.

Two issues then arise as this individual picture increasingly becomes a “facsimile” the real person. First, the lack of control or

transparency regarding who is using our information and how is troublesome. Secondly, there develops a sense of ownership relative to the real person; a sense that we are creating something from which others are profiting or that we may even be harmed by. What this paper is *not* about is identity theft by hackers or overtly malicious thieves.

My assumptions are:

1. Data retention by parties outside the self exists and is increasingly possible and probable and that the exchange/sale of this information occurs.
2. Methods of aggregating and inferring personal profiles or “painting a picture” of that person exist and will continue to evolve.
3. Many current and future technologies we currently enjoy (i.e., will not give up) inherently require some degree of private information to be captured in order to function. Additionally, these devices (online activity, cellular communications, GPS, etc.) will likely become increasingly integrated in both hardware and service provider (repository database). An example would be the iPhone.
4. Individuals have an inherent right to basic forms of personal privacy; Privacy is the notion of being able to control what others come to know about you.<sup>5</sup>

### **The Exchange Concept: Private and Commercial Sector**

Personal data acquisition events occur when an individual exchanges goods, services, or funds with another party, such as health

care, banking, local or online retail sales, cable TV, eBay, etc. The point is, there is a readily perceivable exchange: I engage Verizon to provide a cellular phone and service, and they do so. What gets interesting is the exchange of services for information. Providers such as Facebook or MySpace offer their wares in exchange for simply signing up. Likewise, a Gmail email account is free in exchange for basic information, no more extreme than first and last name, a secondary email, and nation of residence. We'll see that that's not all they are looking for or get. In these cases, the personal information (and the prospect for more) is worth providing what most users would perceive as a service of definable value.

On the surface, there can be some genuine benefits to both parties beyond the initial exchange: better customer service, potential for a discount or rebate program, *pull* versus *push* marketing or production, for example. And even here, there is some degree of equity in the exchange. But what about that information...may it not have other uses?

### **The Valuation of Personal Data**

The issue of retained data can be seen in terms of value. If there were no value attached to the information, it would not be retained, nor would individuals care that it was. Under current market trends and legal structures, especially in the U.S., the trend of retaining and aggregating data will not subside any time soon. For the collectors of information, it has become far too valuable and easy to collect and store. This is happening in two distinct areas: data aggregators/brokers such as Choicepoint and interactive online advertising via the likes of MySpace, Amazon.com, or even relatively evolved specialty online retailers.

Data aggregators collect personal information from businesses, combine it, and then re-sell the packaged information to other businesses or back to the original business. “In insurance, for instance, they aggregate underwriting information from the insurance carriers (which cannot share the data directly due to anti-trust laws) and sell it back to them.”<sup>6</sup> In a more individual-level example, one can go online to the “Swipe” personal information calculator ([www.turbulence.org/Works/swipe/calculator.html](http://www.turbulence.org/Works/swipe/calculator.html)) that has been developed from monetary figures taken from the main nine commercial data warehouses including Accurant, ChoicePoint, ChoiceTrust, DocuSearch, Experian, KnowX, Merlin Data, and Pallorium. If you add up a single request of each of the 46 different personal records, the total valuation is \$277.60.<sup>7</sup> At the other end of the spectrum, Choicepoint’s 2005 total revenue was \$1.1 billion, up 15% from the previous year.<sup>8</sup>

In the case of online advertising, the collection and retention of data is one of the business model underpinnings of companies like Google, described by Economist magazine as “the world’s most valuable online advertising agency disguised as a web-search engine.” And that when Google implemented technology known as AdSense, “a system that goes beyond search-results pages and places “sponsored” (i.e., advertising) links on the web pages of newspapers and other publishers that sign up to be part of Google’s network. Like AdWords these AdSense advertisements are “contextual”—relevant to the web page’s content—and the advertiser pays for them only when a web surfer clicks. Together, AdWords and AdSense produced \$6.1 billion in revenues for Google last year”.<sup>9</sup>

Beyond Google, Amazon.com last year applied for a patent that includes the option of capturing “Education Levels... Genders... Income

Levels...Ages... Geographic Locations... Interests... Races... Ethnicities... Religions... Occupations... Sexual Orientations”<sup>10</sup> when filling out an online order form. More to the point, when inquired about the patent, a company spokeswoman replied that Amazon.com had "no current plans" to implement such a system, but that "Who knows 10 years down the road or five years down the road? It might be good to implement. We want to protect *our* intellectual property."<sup>11</sup>

Or take MySpace, the social networking site: They not only want your personal information to register, but “MySpace.com collects user submitted information such as name, email address, and age ...[and] also collects other profile data including but not limited to: personal interests, gender...education and occupation...This non-personally-identifiable information may be shared with third-parties to provide more relevant services and advertisements to members.”<sup>12</sup> I would point out that while not only does advertising keep the lights on at MySpace, but that they go beyond capturing searches or purchase histories: The major component of their existence hinges on the content that its users upload.

Finally, according to the Interactive Advertising Bureau and PricewaterhouseCoopers, *in the first Quarter of 2007 alone*, internet advertising revenues hit a record \$4.9 billion, a figure up 26% from the same period the previous year.<sup>13</sup> The point is that the retained data will be simply too valuable to be allowed to be deleted or “expire” as the previous authors suggest.

### **The Inequity of Exchange**

Whatever entity is collecting, retaining, and aggregating information on the individual, the data continues to be able to paint a

clearer and clearer picture of that individual as a whole. While the collector of the data is obviously involved in the process, it is still the individual that creates the image. It is as if the collector/aggregators provide the canvas and the paint, but we do the act of painting. Emphasis here is that the “painting” has the potential to increasingly become a facsimile of *us*, and therefore implies some sense of entitlement or ownership. We are creating something of value from which we will not profit. In fact, we fear that we could be harmed by it.

A counter-point to this might be made in that MySpace users can utilize the service to enhance their personal businesses via exposure, and thus achieve greater equity in the exchange. An acquaintance of mine is a perfect example, in terms of her high-end jewelry business: Her presence on MySpace generates not only advertising for her business, but affiliated sales. Likewise it could be raised that the mid-sized retailer that I worked for and utilized on-line (real time networked) Google spreadsheets (a free service) received much in return. I would counter that we are discussing the value of the captured and retained information itself.

If we use GoogleTalk, Google’s free instant messaging service related to Gmail as an example, I suggest that this exchange of information for email service quickly becomes inequitable. There you are, instant messaging away, feeling secure and enjoying a tangible service. Yet according to Google’s privacy policy: “When you use Google Talk, we may record information about your usage, such as when you use Google Talk, the size of your contact list and the contacts you communicate with...You may delete your contacts information or chat histories you have stored in your Gmail account by deleting them through Gmail or by deleting your Gmail account from the Google Accounts page. Because of

the way we maintain this service, such deletion may not be immediate, and residual copies may remain on backup media.<sup>14</sup>

Thus, as you fill up Google's archives with your content, you accordingly build up the value of these archives. The individual retains the same level of services and value that they had at the inception of the relationship, but Google's inventory of profile assets builds and builds. And what of this "picture" of our selves, evolving into clearer focus and higher value? Could it be argued to take on the aspects of published "art" or product? Hence, taking on an aura of something that is approaching the concept of copyright-able?

### **The Lack of Participation and Control: Who Owns Our "Picture"?**

We can all admit and recognize that there is a degree of information about us that is publicly available. It is not so much the information that we willingly (and often routinely) give out or is public record at the local DMV that is alarming. It is the notion of increasingly surreptitious or disingenuous gathering and archiving coupled with active efforts to evolve profiling that becomes disagreeable. As Cassidy, Chae and Courtney summarize in their Legal and Economic Assessment of the Information Externality, where their "externality" is analogous to our "picture": "The externality related to consumer information exists because only business is in the position to collect, compile, and benefit from consumer information...Intuitively our sense of distributive justice suggests that if you create or consume value in society, you should enjoy the benefits or pay the costs." They also point out that this information is "not consumed through use...it is difficult to restrict the use of

information to those who have purchased it...[and] can be resold to subsequent customers.”<sup>15</sup>

I contend that it is the sense of ownership illustrated above, coupled with the loss of control over this representation of the self, post-exchange, where we can look for a solution.

### **If Our Pictures Are so Valuable, Then Don't We Have a Right to Remunerations?**

I suggest:

1. Development of a framework where by individuals may monitor the summary acquisition and distribution/use of their personal information.
2. Build systems that allow individuals to commodify or “monetize” their personal information.

Giving individuals access would allow them to monitor their profile(s) for accuracy and use. It would foster participation with our information post-exchange, as well as open general dialogue regarding how personal information is being used. In short, it would keep things “above board”. While beyond the scope of development here, just as we have Credit Reports, so too could we not have Profile Reports and/or an electronic information portfolio? As for the argument that the retainers would never allow individual access, James Lee, chief marketing officer for Choicepoint surprisingly concedes, “The mantra around here these days is, ‘Why do you have my data, and how am I going to benefit from you using it?’ That's something we're going to devote time and attention to going forward. Transparency is a good thing.”<sup>16</sup>

Allowing individuals to market their personal information would address the inequity of valuation, post-exchange. If the information is of such value, then we should be reimbursed accordingly. As Kevin Maney of the USAToday states: "If the producers - who are, ironically enough, consumers like you and me - could gain enough control to charge for their information, Web sites would be more judicious about asking for it. Simply taking your information would become stealing."<sup>17</sup> That's exactly what Charles Manning, concurs with: "It's time for us as consumers to monetize our data, "Think of this data as real money, because everybody in business does."<sup>18</sup>

## **Conclusion**

Giving individuals methods to not only monitor but manage the use of their personal data can help them regain a sense of involvement and control over their captured information. Adding systems by which they might actually profit or benefit from the proliferating use of this information will increasingly restore equity and fairness to personal data exchanges. In the end, if someone is compiling and selling a biography of your life, you might enjoy being entitled to a percentage of the sales

## Works Cited

1. Mayer-Schönberger, V. (2007). Useful Void: The Art of Forgetting in the Age of Ubiquitous Computing *Faculty Research Working Papers Series: John F. Kennedy School of Government – Harvard University*. Retrieved October 15, 2007 from <http://ksgnotes1.harvard.edu/Research/wpaper.nsf/rwp/RWP07-022>
2. Blanchette, J.F., & Johnson, D.G. (2002, January/February). Data retention and the panoptic society: The social benefits of forgetfulness. [Electronic Version]. *The Information Society* 18(1):1-13.
3. Freeman, Samuel. (2002). *Cambridge Companion to Rawls*. West Nyack, NY, USA: Cambridge University Press.
4. Rose, C. (2004). *Property in Question : Value Transformation in the Global Economy*. (K. Verdery, Ed.). Oxford GBR: Berg Publishers.
5. Deighton, J.A. (2006). The Presentation of Self in the Information Age [Electronic Version]. *Harvard Business School Working Knowledge*.
6. MacDonald, J. (2006, February 21). *How much are your personal details worth?* Retrieved October 25, 2007 from <http://www.bankrate.com/brm/news/pf/20060221b1.asp>
7. *Ibid* 6
8. da Costa, B., Schulte, J., & Singer, B. New Radio Performing Arts, Inc. (n.d.) SWIPE Toolkit. Retrieved October 25, 2007 from <http://turbulence.org/Works/swipe/main.html>
9. (2006, July 6). The ultimate marketing machine. *The Economist*. [Electronic Version]. Retrieved from [http://www.economist.com/business/displaystory.cfm?story\\_id=713890](http://www.economist.com/business/displaystory.cfm?story_id=713890)  
5
10. US Patent and Trademark Office. (2006). *Providing gift clustering functionality to assist a user in ordering multiple items for a recipient*. (United States Patent Application No. 20060178946). Washington, D.C.: Agarwal, A.D.

11. Harris, C. (2006, August 11). Amazon database would put shoppers' intimate details on the line. *The Seattle Post-Intelligencer*. Retrieved October 15, 2007 from <http://seattlepi.nwsourc.com/business>
12. MySpace.com. (2005). Privacy Policy. Retrieved October 27, 2007 from <http://www.myspace.com/index.cfm?fuseaction=misc.privacy>
13. Interactive Advertising Bureau & PricewaterhouseCoopers LLP. (2007, June 6). *Internet Advertising Revenues Soar Again, Near \$5 Billion in Q1 07*. BusinessWire. Retrieved October 28, 2007 from [http://www.businesswire.com/portal/site/digg/index.jsp?ndmViewId=news\\_view&newsId=20070606005799&newsLang=en](http://www.businesswire.com/portal/site/digg/index.jsp?ndmViewId=news_view&newsId=20070606005799&newsLang=en)
14. Google. (2005, October 14). *Google Privacy Policy*. Retrieved October 20, 2007 from <http://www.google.com/privacypolicy.html#information>
15. Cassidy, C., Chae, B., & Courtney, J. (2004). Ethical Management of Consumer Information: Solving the Problem of Informational Externality Using the Coasian Approach. In Freeman, L. (Ed.), *Information Ethics: Privacy and Intellectual Property*. (120-138). Hershey, PA, USA: Information Science Publishing.
16. *Ibid* 6
17. Maney, K. (2000, May 31). Personal data for sale. *USA TODAY*. Retrieved October 27, 2007 from <http://www.usatoday.com/tech/>
18. *Ibid* 17

