Data Mining the Kids: Surveillance and Market Research Strategies in Children's Online Games

Grace Chung
Sara M. Grimes
Simon Fraser University

Abstract: This paper explores privacy issues in relation to the growing prominence of marketing research and data mining in websites for children. Whereas increasing protection is given to individuals’ personal information, little attention is paid to information that is aggregated, electronically scanned, and sorted—despite the fact that aggregate information is often highly valued by the marketing industry. The authors review current trends in Internet market research, data mining techniques, policy initiatives, and the contents of some of the most highly frequented children’s game sites. The paper demonstrates how common data mining practices constitute a threat to children’s emerging rights online.

Résumé : Cet article explore la question de la vie privée des enfants et la proéminence croissante de la recherche marketing et le « data mining » dans le développement et gestion des sites Web pour enfants. De plus en plus, l’incorporation des technologies «data mining» dans les sites Web résultent en un fusionnement de contenu et commerce, ce qui élève un nombre de questions concernant les droits des enfants et les responsabilités des chercheurs. Les auteurs concluent que les droits naissants des enfants en ligne sont à risque dans ce contexte commercial de l’Internet.

Keywords: Children’s media; Content analysis; Cultural studies; Ethics; Internet/IP/WWW; Policy

Introduction

Over the past decade, the widespread adoption of online tools and technologies by children has become the topic of much discussion within both academic and public spheres. As children and youth have expanded their presence on the Internet (Buckingham, 2003; Holloway & Valentine, 2003; Jenkins, 2001), they have been depicted by some as adopting increasingly participatory roles in the creation of online content through their contribution to online environments,
games, and communities. What this literature overlooks, however, is that the many popular children’s sites are often commercially owned and operated, responding primarily to advertiser demands and other corporate interests (Kline, Dyer-Whit-eford, & de Peuter, 2003; Montgomery, 2000; Shade, 2004).

As traditional barriers between “content and commerce” disappear, the result is the emergence of what Montgomery (2000) terms a “children’s digital media culture,” wherein new levels of intimacy are built between marketers and children. This relationship is clearly illustrated within popular branded children’s online game communities, such as Neopets.com, EverythingGirl.com, and Postopia.com. While these sites provide young users with hours of entertainment and endless online play opportunities, they simultaneously engage them in data mining activities that transform children’s play into a way of gathering information. This phenomenon, and the corporate mechanisms that drive it, is reflective of a larger trend in online-gaming conventions—one that increasingly incorporates marketing-research strategies into the design and operation of online entertainment sites and virtual communities.

Data mining technologies create detailed demographic and behavioural profiles of children and young people online. This information is then used to document contemporary trends in youth markets—a research practice often referred to, by marketers, as “cool hunting.” Data mining is the process of gathering “aggregate” rather than “personal” information. Whereas there is increasing protection of individuals’ personal information online—that is, information that lends itself to the unique identification of one individual from another, (including name, age, social insurance number, address, date of birth)—little attention or protection has thus far been given to information that is aggregated, electronically scanned, and sorted. Yet aggregate data is a highly useful and sought-after commodity for marketers, who compile extensive databases containing information on users’ online browsing habits and consumer preferences, for example, from which they are then able to extract meaningful patterns and relationships. This process of yielding consumer information for commercial forecasting raises a number of ethical concerns, including questions of privacy and intellectual property ownership in relation to non-personally identifiable information and content, resulting in what David Lyon terms “social sorting” or what Danna and Gandy identify as “weblining.” Social sorting, as Lyon explains, codes and classifies aggregate groups. This influences, manages, and affects the choices and life chances of individuals (Lyon, 2002). “Weblining” refers to a process by which consumers may be denied certain goods and services based on their categorical online profiles (Danna & Gandy, 2002).

This paper explores privacy issues in relation to the growing prominence of marketing research and data mining on websites for children. The following discussion provides a preliminary examination of marketers’ usage of data mining techniques to collect and interpret various types of information about children within the context of online game sites. A brief overview of the current debates surrounding children’s relationships with new-media technologies outlines the
sociocultural context within which these processes are seen to operate. A number of contemporary industry trends within market research and more general forms of online surveillance are examined, alongside an overview of recently implemented Canadian and U.S. policies that address Internet privacy issues, to explore the ethical dimensions of applying data mining strategies to children online. In addition, our discussion draws upon the findings of a year-long comparative multi-case study of seventeen popular children's gaming websites, including the online game communities Neopets.com, EverythingGirl.com, and Postopia.com. Using Yin's (1984) model for multi-case, embedded case-study design, the methodology for this study combines both qualitative and quantitative analysis of content with an historical overview of the corporate activities, holdings, reports, and partnerships of the games' owners and operators.

The children's digital playground

The rise in the importance and visibility of the child consumer within North-American markets since the mid-twentieth century has led to the creation of a seemingly endless barrage of new product lines, services, and technologies directed toward children and teens. Within the context of the media and cultural industries, this transformation in children's roles and functions has translated into the development of an equally specialized (and increasingly profitable) niche audience and consumer market. With the emergence of every new media technology, children's culture becomes increasingly inundated by the mass media and the sweeping marketing efforts of toy companies and other child-focused consumer-goods industries (Kapur, 1999; Kline, 1995; Montgomery, 2000). Concurrently, children's spending power also has increased significantly. Sutherland & Thompson (2001) report that children's spending has "doubled during each decade of the 1960s, 1970s and 1980s and has tripled in the 1990s" (p. 79). Recent U.S. estimates place the value of the children's market at around $115 billion ($1.8 billion for the Canadian "tween" market—a term used to describe children between the ages of 9 and 12—alone), including both the money children spend themselves and the influence they exert over family purchases (Sutherland & Thompson, 2001). Accordingly, advertisements for adult products, such as food, electronics, furniture, and hotel chains, are increasingly targeted toward children (Kapur, 1999), who are often celebrated within media and marketing discourse as "cynical" and "sophisticated" consumers. This discourse has contributed to a growing acceptance of the notion of the "efficient, autonomous, and flexible child" (Hultqvist & Dahlberg, 2001, p. 9) that is currently found within political, pedagogical, and social thought.

The image of the "affluent child" is often accompanied by the image of the "media-savvy" or media-saturated child (Holloway & Valentine, 2003), as studies consistently reveal that children are spending increasing amounts of their time with the media. In 1999, for example, the Kaiser Family Foundation reported that American children were spending an average of 5 hours and 29 minutes a day, seven days a week, using media for recreation (Edwards, 1999). In addition, much of this media usage was reportedly spent alone and unsupervised: 53% of children
surveyed had a television in their bedrooms, 70% had a radio in their bedrooms, 64% had a tape player, and 16% had a computer (Edwards, 1999). The notion of the “cyber-kid” has also become prevalent in recent years, due in part to children’s early adoption of information-communication technologies (ICTs) and the large amount of time they spend with them. A recent study conducted by Nielsen/Netratings (2003) reveals that children and teens account for one out of every five (21%) Internet users in the U.S., totalling more than 27 million (12 million between the ages of 2 and 11 years and 14.9 million aged 12-17). Meanwhile, in Canada, single-family households with children under the age of 18 had the highest rate of home Internet use, constituting approximately 73% of the estimated 6.7 million Canadian households with home Internet access in 2003 (Statistics Canada, 2004). Since families with children remain among the fastest-growing demographics of Internet users (Montgomery, 2000), the number of children online continues to multiply. Children are also using digital-media technologies at an increasingly younger age. As Rideout, Vandewater, and Wartella (2004) describe, “Nearly half (48%) of all children six and under have used a computer, and more than one in four (30%) have played video games” (p. 4).

As a result of the vast amount of time children and young people are reportedly spending with the media, they are often framed within the popular press and marketing literature as media-“savvy” experts. This perspective is also found throughout media and policy discourses surrounding children’s usage of ICTs (Livingstone, 2004), as well as in numerous studies arguing that access to new-media technologies empowers children and youth through the creation of online content and contribution to online environments, games, and communities (Buckingham, 2003; Holloway & Valentine, 2003; Turow & Kavanaugh, 2003). The idea being promoted here is that children have a seemingly natural ability to master digital technologies and navigate the information society. Computers and other ICTs are furthermore championed, as Seiter (2004) suggests, as the “ultimate fulfillment of adult desires to see play turned into a purposeful end, to use play for progress and child development—all this without children noticing the beneficial effects” (p. 93). Studies show that parents see computers as predominantly educational—including a recent survey in which 72% of parents reported that computer use “mostly helps” in their children’s learning (Rideout et al., 2004). Perceptions of children’s adeptness with ICTs are inflated when contrasted with the challenges and difficulties many parents face in adopting these same technologies. Kapur (1999) writes:

The “knowledge gap” between adults and children is closed . . . because of the rapid disappearance of the world adults knew and learned about as children. . . . New technologies, such as computers, which are familiar to children but new to adults, place the child ahead of the adult in achieving a working knowledge of the world. (p.129)

Contemporary notions and representations of the Western child—as sophisticated consumers and media-savvy cyber-kids—thus support the notion that children have a high level of competency and skill when it comes to understanding
cultural discourses and manipulating technologies. However, the exact extent and nature of this knowledge has yet to be clearly determined.

Recent studies of children’s economic socialization, for instance, show that “although today’s children and adolescents have the spending power to utilize their consumer skills, they still often lack the maturity to think carefully about buying decisions” (Valkenburg & Cantor, 2002, p. 211). Furthermore, little distinction is currently made between media- or peer-informed consumer decisions and more informed choices or “thoughtful consumer decisions” (Valkenburg & Cantor, 2002, p. 211) within either consumer-socialization research or marketing rhetoric. In addition, and in contrast with the ideals put forth by the notion of the “cyber-kid,” emergent studies suggest that children’s knowledge of the processes and issues associated with ICT usage might actually be quite superficial and limited. For example, in a 16-week study of children’s Internet use within a public-library setting, Sandvig (2000) found only nine requests for privacy policies of any kind among the 203,647 page requests submitted by children during that time period. More recently, Shade, Porter, and Sanchez Santiago (2004) report that young children have difficulty understanding questions about privacy, know very little about common Internet business practices such as sending “cookies” to track users, and often do not fully comprehend why personal information should not be divulged online. Turow (2001) reports that children are more likely than adults to give out sensitive information, particularly in exchange for a free gift or reward, and that 46% of parents are unaware that websites gather information on users without their knowing it. These findings support a growing body of academic research demonstrating the limitations of children’s understanding of Internet (especially corporate-website) processes (Seiter, 2004; Shade, Porter, & Sanchez Santiago, 2004), as well as children’s overall lack of critical sophistication when it comes to new media (Kline, 2001; Livingstone, 2003). They also suggest that private-industry standards for securing informed consent from child users to the terms, concepts, and processes at work within children’s websites and Internet applications are inadequate at best. From the lack of attention children give to website privacy policies to their limited understanding of key legal and economic concepts, the findings of these studies illustrate the need for further inquiry into children’s actual (as opposed to assumed) cyber-literacy, as well as the need to identify the exact nature and implications of the exchanges and processes in which children are engaged online.

This is not to say that new forms of interactive media and the ready accessibility of online tools and software—for the production of digital film, animation, games, webpages, and weblogs—have not allowed children to adopt an increasingly participatory role in the creation and sharing of cultural artifacts. We contend that while the proposition that young people actively engage with the mass media and to a degree forge it in their own image is a sound one, it is also one that is only ever partially realized in today’s commercialized digital landscape. Ultimately, the parameters within which young people are able to manipulate and contribute to online culture are often set within a corporate framework
that is inevitably constructed first and foremost by a desire to build profits and
nurture brand identity. Popular children’s websites increasingly are characterized
by the substantive role played by marketing research in the development and
maintenance of online activities and communities that are directed toward gath-
ering user information and transmitting advertising messages. As barriers to
accessibility and participation dissolve, distinctions between private and public
space conventionally used to define and delineate childhood have also disinte-
grated (Cook, 2001). New media and other digital technologies— together with
the forces of global capitalism—have made childhood not only “inseparable from
media use,” but also intimately tied to emerging forms of media and audience sur-
veillance (Cook, 2001, p. 82). This process has “intensified with the newer, more
elaborate tracking techniques and organization of data made available through
computers” (Kapur, 1999, p. 126). New technologies and tools, such as those for
data mining, are changing the very nature of consumer research through the appli-
cation of increasingly unobtrusive, yet highly intrusive, ethnographic-research
methods and the collection of enormous amounts of data on consumer habits,
preferences, and daily activities. These practices are part of a growing trend in
consumer research that Russakoff (1999) has labelled “cool hunting,” a practice in
which marketers “get kids talking about their taste-worlds” (Quart, 2003, p. 42).
Recent market-research literature, such as that by Lindström (2003), praises
under-the-radar techniques such as “community exploration, peer-to-peer mar-
ketin g and viral marketing,” as the “key ingredients” to almost every successful
“tween” brand on the market today (p. 138).

“Cool hunting” in the children’s digital playground manifests itself in various
forms, foremost among which are the innumer able commercial game sites that
offer children compelling and entertaining online spaces in which to build com-
unities, create content, and play games. Concurrently, however, these same sites
award “points,” virtual currency, or similar incentives to players who fill out mar-
keting surveys, visit sponsored sites, and participate in simple point-and-click
“advergames.” Although many of the more overt marketing-research methods
have fallen under the mounting scrutiny of consumer watchdogs and child-advocacy
groups—as well as the U.S. government’s Children’s Online Privacy Protec-
tion Act (COPPA), implemented in 2000—more subtle approaches to data mining
remain for the most part unregulated and unrestricted. For example, no protection
currently exists for information that is collected and stored in aggregate form,
though detailed studies of whole demographics and interest groups is often what
marketers and advertisers value most (Lindström, 2003; Smith & Clurman, 1997;
Sutherland & Thompson, 2001). More and more, youth-market research is the
“ongoing collection of personal information and tracking of on-line behaviour”
(Montgomery, 2000), which is then stored and manipulated in aggregate form.
Despite this trend, much of the policy on children’s privacy in relation to mar-
keters has been limited to the prohibition of any direct solicitation of personally
identifiable information from children under the age of 13 unless parental consent
is granted.
Encompassed in the notion of “dataveillance,” which involves the “automated monitoring through computer-readable data rather than through physical observation” (Clarke, 1994), data mining uses a range of virtual tools to draw upon raw material from any number of online or offline activities or transactions, using data algorithms to sift through large quantities of information and aid in the knowledge-discovery process (Danna & Gandy, 2002). Originally applied in the research of artificial intelligence, automated mining technologies allow for the efficient discovery of otherwise non-obvious, previously unknown information, facts, and/or relationships. The practice of data mining is enhanced through the process of extracting “operational” data and transforming it into “informational” data, then loading it into a central data store or data warehouse. Data mining techniques use specific algorithms to establish correlations or patterns in data on which decisions about future courses of action can be based. In this regard, “the raw material is the business data, and the data algorithm is the excavator, sifting through the vast quantities of raw data looking for valuable nuggets of business information” (Cavoukian, 1998, p. 4). Companies are thus able to use sophisticated user-tracking services, data-collection and data-storage technologies, and complex database algorithms to create detailed consumer profiles and map behavioural trends.

In further characterizing the online commercial environment, Danna and Gandy (2002) write, “more information is being captured in Web server logs. Sophisticated analytic and data mining software tools enable firms to use the data contained in these logs to develop and implement a complex relationship management strategy” (p. 373). The process is particularly unique, Barney (2000) argues, as it enables the extraction of unintended or unexpected “patterns and relationships, that the user may have never considered looking for” (p. 226). As he explains, data mining algorithms search through raw data, identifying trends, drawing associations, locating sequences and clusters, and constructing generalizations and categories. When networked to other databases and data warehouses, a truly staggering amount of consumer information can be gathered and processed in increasingly useful ways. As Lyon (2001) describes it, “Database marketing is now heavily involved in... [a] ‘world wide web of surveillance’” (p. 101) or “cybersurveillance” (p. 145).

The privacy of personal information has thus become a pressing concern for users of the online environment. In 1998, a survey carried out by Georgia Tech showed how more than half of the participants had concerns about Internet privacy and security. However, only a small proportion of adult Internet users pay attention to the contents of website privacy policies and terms-of-service contracts, let alone fully understand their legal implications. In the U.S., as Turow (2003) describes, the majority of adult Internet users “misunderstand the very purpose of privacy policies” and erroneously believe that the mere inclusion of a privacy policy indicates that the website “will not share their personal information with other websites or companies” (p. 3). In the case of young children, who are even more likely than adults to divulge personal information in exchange for a
reward or prize, for example, the question of whether informed consent is secured and whether user rights are protected within the online context becomes even more doubtful and problematic. These issues have been recently brought to the forefront of policy agendas in the global arena (e.g., by the UN Convention on the Rights of the Child or the U.S. Federal Trade Commission). We will now focus briefly on two recently implemented frameworks—namely, the Children’s Online Privacy Protection Act (COPPA) in the U.S. and the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada—that reflect how government agencies have addressed privacy issues, combining consideration for children’s rights with facilitation of the online economic activities of private industry.2

In response to growing public debate over the commercialization of children’s online spaces, COPPA was passed by the U.S. Federal Trade Commission (FTC) in 1998, taking effect in April 2000. COPPA established new laws regulating the relationship between marketers and children online, requiring that commercial websites aimed at children under the age of 13 conform to a variety of guidelines aimed at regulating children’s interactions with marketers online. Included in this new set of laws is the requirement that children’s websites (both those directly aimed at children and those that knowingly gather information from users under the age of 13): (a) give parents notice about their data-collection activities; (b) obtain verifiable parental consent prior to collecting information from children; and, (c) provide parents with access to any information collected from their children, as well as the opportunity to discontinue any further uses of the data collected (U.S Federal Trade Commission, 1998; Montgomery, 2000). One of the most important shortcomings of the act is its narrow definition of privacy and what constitutes personal information. For instance, the COPPA laws apply only to personal information from a child in identifiable form and do not take into account data collected and stored in aggregate form.

In response to Canada’s challenge finding a balance between the private sector’s progressions in information technology and data management, the Parliament of Canada passed a law in 2000 to outline national rules for private-sector collection, use, and disclosure of personal information. Known as PIPEDA, this legislative framework for data protection encompasses internationally agreed-upon principles whereby an organization in the private sector

must be accountable for all personal information in its possession; should identify the purposes for which the information is processed at or before the time of collection; should only collect personal information with the knowledge and consent of the individual; should be open about its policies and practices and maintain no secret information system . . . (Bennett & Grant, 1999, p.6).

Full implementation of the legislation came into effect at the start of January 2004. Yet as Barney (2000) suggests, Canada’s privacy policy “does not even contemplate that the digital collection, use, and exchange of personal information is something that perhaps should be fundamentally limited” (p. 229). It merely requires that companies state their purpose and obtain some form of consent
before collecting data from consumers. Furthermore, the only mention of minors in PIPEDA is found in Clause 4.3 Principle 3, which states, “Seeking consent may be impossible or inappropriate when the individual is a minor, seriously ill, or mentally incapacitated” (Department of Justice, 2000).

A comparative analysis of these two policies reveals a set of complementary strengths and limitations. While the U.S. has shown exemplary progressiveness and forward thinking in terms of formulating a legislative framework to recognize and pay attention to the human rights of children (in this case, privacy rights), the Canadian legislation (PIPEDA) recognizes emerging problems and dimensions for consumers’ rights within the context of online economic activities. COPPA fails to adequately define and address the full spectrum of commercial transactions that are taking place within children’s online environments (through a very limited approach to privacy and what constitutes personal information). PIPEDA, meanwhile, fails to give sufficient consideration to children who are perhaps unaware or unable to enter into informed legal contracts. A more comprehensive approach would combine the aforementioned strengths of both acts, while omitting the oversights that have thus far limited the potential effectiveness and impact these frameworks might otherwise possess. In other words, these existing policies fail in different areas to offer adequate protection and promotion of children’s rights online. For children’s digital play spaces to truly promote and protect the rights of the child within economic contexts, an amalgamation of the principles of these two countries’ legislative responses to children would be required.

**Data mining and ethical dimensions to privacy**

To better understand the processes at work within commercially owned children’s websites, we draw upon the results of a year-long, multi-case, embedded case study of seventeen online game sites popular among contemporary children and youth (Grimes, 2005). Cases were selected based on rankings measuring the online destinations most frequented by “children” (aged 2-12 years) and the more general category of “youth” (users under 18 years), as reported by Internet audience-research firms Nielsen/Netratings (2002) and Hitwise (Greenspan, 2003). Not surprisingly, the case-study findings support previous work by Kapur (1999), Montgomery (2000), and Cook (2001), among others, who identify the Internet as the site of an increasing amalgamation between children’s culture and marketing initiatives. Almost all of the children’s sites analyzed incorporated advertisements and market-research strategies into child-oriented games and entertainment-driven content. In the most extreme case included in the sample, Neopets.com, a site that claims over 22 million young members worldwide (39% of whom are under the age of twelve) (“Neopets Press Kit,” Neopets, Inc, n.d.), aggregate information about player demographics and online behaviour was packaged and sold to clients and trade publications within children-focused industries in the form of demographic studies and youth-trend reports. In her own study of this particular website, Seiter (2004) contends, “Instead of selling a media product itself, Neopets is selling information about the children and young adults who are its fans” (p. 98). In another case, that of the Barbie/Mattel webportal Everything-
To participate in these sites, users must first sign up for a free membership by disclosing a limited amount of personal information and creating a unique online identity. Nearly all of the sites reviewed collected some form of personally identifiable information from the player—e-mail address, date of birth, and gender being the most common, though some also asked for the user's name and at least some components of their home address, such as state/province, country, or zip/postal code. This condition of play is in accordance with the classic virtual-community requirement articulated by Kollock (1996), namely, that trust be built or facilitated among users via persistent member identities. It also allows the companies to follow and record users' behaviours and preferences as they move through the sites, through both covert tracking software and more obvious forms of market research that are often firmly integrated into the site content. Neopets.com, for example, awards users who complete surveys on a wide variety of topics and product preferences a large number of “Neopoints” (the currency required to purchase food and items for the virtual pets at the centre of the Neopets game), and attracts up to 8,000 young respondents a day. The EverythingGirl.com site rewards players who complete a weekly selection of “featured games”—most of which consist of little more than product-preference surveys—with “Pippa-Points” that users can apply to the purchase of “Charms” (all of which feature one of Mattel’s branded characters, such as Barbie or Polly Pocket) for a virtual “charm bracelet.” In another instance, Kraft Foods’ Postopia.com site encourages players to earn “Postokens” by submitting a “code” found on the inside of specially marked boxes of child-oriented Post cereals (including Fruity Pebbles, Alpha-Bits, and Oreo O’s). By integrating research protocols with features of the game play, as in the case of Neopets.com and EverythingGirl.com, or with consumer incentives to buy the very products being promoted on the site, as in the case of Postopia.com, the sites motivate users to provide a continuous supply of personal information and other useful market-research data.

Facilitating the corporate appropriation and commodification of children's submissions and responses to these sites are end-user licence agreements (EULAs) or terms-of-service (TOS) contracts. Participation in the vast majority of the websites reviewed required young users (and/or their parents, though little is in place to ensure that parental authorization is legitimate) to first agree to sweeping terms-of-service contracts that void them of any rights over their contributions or communications (Grimes & Shade, 2005). For example, the terms-and-conditions contract found on the EverythingGirl.com website provides a detailed overview of the various ways a child’s contributions may be used by the site owners once copyrighted:
You grant Mattel a non-exclusive, royalty-free, perpetual, irrevocable, and sub-licensable right and license to reproduce, distribute, publish, transmit, modify, adapt, translate, display, distribute [sic], sell, license, publicly perform, prepare derivative works based upon, and otherwise use or exploit Your Submissions throughout the world in any and all media. (Mattel Inc, 2003)

In other words, these kinds of terms-of-use contracts ensure that the valuable user data mined from user responses, submissions, and online interactions and behaviours becomes the exclusive and unlimited property of the website owners. This condition of use has proven quite lucrative for companies like Neopets, Inc., whose revenue is in part derived from the aggregate demographic data sold to marketers and members of child-oriented industries in the form of youth-market-research reports (Grimes & Shade, 2005). With an executive team drawn from the online marketing-research industry, including alumni from OpinionSurveys.com and Netmagic, Neopets Inc.’s annual “Youth Pulse” research reports have become a staple resource for Fortune 1000 companies targeting children and youth, as well as industry news sources such as Advertising Age magazine (Grimes & Shade, 2005). The company boasts a large portfolio of high-profile clients, including Mattel, Lego, General Mills, McDonald’s, and Disney, as well as a number of international advertising agencies (“Neopets Press Kit,” Neopets, Inc., n.d.).

In this respect, children’s sites such as Neopets.com represent a new sustainable revenue model for online communities—that of transforming community members into a stable resource for market-research initiatives—in addition to profiting from more established revenue models. It is in this sense that the users themselves are transformed into commodities. Data mining technologies, as Mosco (2004) writes, are used “to refine the process of delivering audiences of . . . computer users, to advertisers. Companies can package and repackage customers in forms that specifically reflect both their actual purchases and their demographic characteristics” (p. 158). In this way, Internet users are reduced to mere audience commodities (Smythe, 1981), as the use value of users’ online experiences and relationships are made subordinate to the exchange value of packaged user-trend reports and data-mined demographic profiles. The users thus become alienated from the digital products of their online interactions and activities, which are aggregated and reconstituted in commodity form as the intellectual property of marketers and website operators.

The increasing use of data mining techniques prominently highlights a new set of ethical dilemmas for the privacy debate. As far back as the 1980s, advances in data-matching and information-sharing across public- and private-sector organizations facilitated the intersection of commercial and administrative domains. The fallout has been that a growing number of interlocking institutions and agencies have come to possess the ability to assemble data images which do not simply follow, but are actually capable of preceding, embodied persons (Stalder, 2002). The formation of and increasing reliance on data images to execute digitally based discriminations present a serious threat to those from whom data is originally derived, as well as for those to whom the images and abstraction are applied. Danna and Gandy point out that there is a margin of error in this process. Within
the context of socially sorting individuals based on their data representations, “given the room for error in data, it is unlikely that a recommendation based on a predication will precisely capture the needs of an individual consumer. Data stored in a record will never be able to truly represent a complex autonomous individual” (Danna & Gandy, 2002, p. 379). We believe that the implications of surveillance practices made possible through networked information and communication technologies simultaneously problematize and call for a continued discourse of definitions of privacy that take affirmations to embodied persons as their primary object of identification. The use of individual data in aggregate form can be seen in the end products of many “cool hunting” initiatives, including Neopets.com’s popular “Youth Pulse” reports (Grimes & Shade, 2005).

At the heart of the privacy debate persists the presumption that a direct connection between embodied individuals and their data can be discerned, implicating as a natural consequence the importance of consent in the formulation of privacy policy or norms (Vedder, 1999). Within Vedder’s formulation, the argument is that an individual person should have the right to consent to how data is abstracted from the digital traces of their activities. This again suggests that the manipulation of such information in the absence of consent constitutes an infringement of privacy norms. As Vedder points out, the flaw in the practices of many websites engaged in data mining activities is that they exclude the social act of granting consent to the use of personal data. Among the many examples where these practices are applied, including EverythingGirl.com, consent is implied, and incentives are offered in exchange for the personal information and product preferences of its young users. Although these sites describe much of the information collected as “non-personally identifiable,” data mining software operates on the basis of open-ended queries, abstracting individuated information into aggregate forms. Open-ended queries are used because data mining functions on the basis of “discovery,” where pattern-matching and other algorithms seek out relationships in data that were not necessarily anticipated prior to data manipulation (Tavani, 1999). Consequently, data mining is to be distinguished from more traditional patterns of information retrieval, in that it is based on implicit searching techniques that use non-predictive queries, serving to construct new and uncertain aggregated images from data warehouses.

An important distinction must be made between the domains of public and personal information. Whereas personal information is commonly understood as corresponding to the intimate sphere—one’s personal data, including birth date, gender, and postal code, for example—public information is not associated with any such degree of intimacy. However, there needs to be a further distinction, between “personal,” or “intimate,” privacy and information privacy. Personal or intimate privacy pertains to one’s ability to protect against violation or exploitation of information sought prior to an intervention. In contrast, information privacy is based on data gained from pattern-matching algorithms that function to develop data images in the first place. In the business practices of at least one of the cases reviewed, Neopets.com, the company was found to collect, process, and
sell its users’ data in aggregate form as “Youth Pulse” reports (Grimes & Shade, 2005). While this particular practice was not evident in the other cases reviewed, including EverythingGirl.com and Postopia.com, the incorporation of “product preference” games and the community-development tools prevalent among these sites nonetheless points to a similar set of processes at work, albeit at a seemingly internal level. Advertisers and corporations are thus able to construct highly targeted marketing campaigns that use the players’ own personal information and behaviours to guide in the development and design of marketing initiatives, as well as to drive new forms of digital discrimination. The discussion will now turn to the wider implications of these processes and how digital discrimination can result in both weblining and the social sorting of individuals.

Implications and tradeoffs to site access
The delineation between types of information and the various forms of privacy is critical to identifying whether data has the potential to become information that carries wider implications such as the social sorting of individuals, which leads to digital discrimination. Here we think it is important to consider the caveats of Moor (1997) in his discussion of normative privacy rights. “Normative privacy rights” describe a type of privacy that “can be lost but not violated,” because there are no specific “norms—conventional, legal, or ethical— which proscribes one’s right to be protected” (p. 30). Moor claims that an individual “has normative privacy in a situation with regard to others if and only if in that situation the individual . . . is normatively protected from intrusion, interference, and information access by others” (p. 30). Although not all data mining practices violate normative privacy rights, certain aspects of the practice need to be further analyzed by other researchers. Moor (1990), for instance, argues that data mining activities result in the alienation of information from the context from which it was mined. When companies mine data to produce profiles of their users to facilitate and expedite the online experience, for example, the context would be evaluated positively by most standards. Yet when the same data is mined for purposes of consumer research and marketing, perhaps by an Internet page that releases or sells the personal information of a user to an external market-research company, the context or situation changes in a practical and an ethical sense. The point is that what counts in the formation of normative privacy guidelines is the situation or context, not the specific information being mined. On Neopets.com, for instance, permission to participate in opinion surveys and divulge information to Neopets.com sponsors is included as part of the parental consent form (which parents of participants under the age of 13 must sign and send in to the site before the child is allowed to participate in the more interactive components of the site), under the guise that these activities “help keep Neopets free for everyone” (“Neopets Parental Consent Form,” Neopets, Inc., n.d.). Surprisingly, many parents willingly agree to these conditions and grant consent on behalf of their child. As Rodgers (2004) reports, “The company gets between 400 and 600 such consent forms faxed to its offices daily and more come through a P.O. box.” What results from this collection and use of personal data by marketers is the clustering of pop-
ulations according to geodemographic type. This is where the social sorting occurs, where “using such clusters in conjunction with postal [zip] codes, marketers sift and sort populations according to their spending patterns, then treat different clusters accordingly” (Lyon, 2003, p.14). The differential treatment of service is known as “digital redlining” or “weblining.” In these instances, “customers are classified according to their relative worth” (Lyon, 2003, p.14).

Considering the trends toward increased data migration these insights prove instructive in understanding the changing nature of the privacy debate. As is often the case, individuals may grant consent for their data to be mined for one purpose and in one context, but when this data finds its way into other mining operations, the context changes. Since much of the activity surrounding advocacy for privacy protection still congregates along the axis of personal-data protection, it is problematic that many data mining applications, which function to render visible new categories of individuals, cannot be anticipated by any one individual. The expansion of searchable databases and the development of sophisticated networked infrastructures change the context in which different solutions are needed. For these reasons, affirmations to privacy norms and personal-data protection are not capable of ensuring the very principles to which they aspire. The issues of informed consent and the user's understanding of the concepts involved in personal-data transactions are particularly significant when the user group in question consists of minors (Grimes, in press). Given the sophistication of corporate data mining capabilities, the assumption that children have a sufficient level of understanding and experience of concepts such as privacy, personal information, and intellectual property ownership to consent to data mining activities is highly unrealistic. Despite the pervasiveness of the myth of the savvy “cyber-child” discussed earlier, it remains questionable whether it is reasonable to expect children to have the skills and knowledge required to understand the concepts and implications of many of the clauses and items found in privacy policies and EULAs—assuming, of course, that they are inclined and able to read through these lengthy and difficult texts in the first place. It is also unclear whether current contract laws in Canada, the U.S., and elsewhere even allow corporations to enter minors (either directly or through parental consent) into legal agreements of this kind. Nonetheless, children and their unique legal status as minors are granted very little special consideration in terms of how privacy policies and EULAs are currently formulated and applied (Grimes, in press). The failure of these sites to address children's special legal status within contract law is particularly problematic, as often minors' contracts (especially those for goods or services not deemed to be necessities) are considered void or at the very least voidable by Canadian and U.S. courts.

Data mining practices thus present a paradoxical situation. While the abstraction of personal data into aggregate profiles tends to mitigate claims to the protection of personal data, those aggregate profiles may be applied against the persons from whom the data was originally abstracted. In one context, the identifying categorical characteristics found in user profiles (e.g., one's birth date and postal
code) may serve to facilitate an efficient and accurate online surfing experience, but in another context, identifying characteristics carry the potential to influence what Lyon (2003) refers to as the “social sorting” of individuals. In this instance, “the resulting classifications are designed to influence and to manage populations and persons thus directly and indirectly affecting the choices and chances of data subjects. The gates and barriers that contain, channel, and sort populations and persons have become virtual” (Lyon, 2003, p.13). These processes and patterns are not unique to electronic information exchange, but data-warehousing and data mining techniques have greatly enhanced the scope to which such procedures apply.

Data images and private-sector profiling have material consequences for embodied individuals. Seeking to accommodate the categorical nature of contemporary surveillance it should not be forgotten that data images are derived from the traces left by embodied individuals. For this reason, a re-configured conceptual approach to privacy is required that can simultaneously account for context and personhood. What is needed is a conception of privacy that, in the words of Lyon (2002), considers personhood as central to its operation but that cannot be reduced to the individualistic level. In this vein, Vedder (1999) advocates the notion of “categorical privacy” as a set of normative principles or values that is irreducible to the level of individual privacy but that remains cognizant of the “individual constitution” of social sorting. Categorical privacy takes the protection of the individual as a central concern, but it emphasizes the generalized properties of categorical images and how they present serious threats to the lives and life changes of embodied persons. Of even more importance is the contention that, while remaining in the moral-legalistic framework of privacy, the shift in discourse from individual to categorical privacy signals a wider connection with normative principles such as social justice, equality, and fairness. Indeed it was these social attributes that Lyon (2002) had in mind when he called for clarity on what constitutes human dignity and social justice within categories and codes gleaned through contemporary surveillance systems.

A first step toward a more comprehensive safeguarding of human rights issues online is a serious interrogation of two interrelated concepts: fairness and awareness. Those who wish to use the services of any of the websites mentioned are not permitted to opt out of the data-collection system without being denied access to the site’s activities and features. Furthermore, it is unlikely that users, especially children and young teens, are actually aware of the fact that their data is being compiled on such a grand basis. Studies into young people’s online behaviours suggest that children hardly ever access website privacy policies (Sandvig, 2003), which are at any rate often laden with legal and technical jargon that is difficult for even adult readers to fully comprehend. The contentious nature of EULAs, which claim sweeping and irrevocable ownership over all that is said and done within the sites, further problematizes these issues, in that the ambiguous and seemingly ubiquitous reach of “terms of use” contracts have thus far allowed website operators to comply to existing children’s privacy laws (where applicable)
on the one hand, yet to accumulate vast databases of richly detailed user information on the other. Though social conceptualizations of intellectual property, privacy, and personhood are being challenged by adult Internet users in various arenas, these same debates are rarely applied to children’s online spaces. While it may be more difficult to imagine knowledge of children’s play habits and social interactions as a highly valuable commodity, the reality of the $115-billion children’s consumer market and the massive revenues collected by children’s marketers prove otherwise. Awareness of the intermediary role data mining and market research play in the relationship that exists between children playing online and the continued growth of children’s commercialized entertainment is crucial to understanding the consequences of the processes that unfold within the children’s digital playground.

**Conclusion**

In reviewing the ethical dimensions of data mining practices found in children’s online game communities, it becomes apparent that if children’s rights online are to be truly promoted and protected, a fundamental shift must first take place in the way that operators of websites for children adapt to COPPA in the U.S. and PIPEDA in Canada. While privacy policies and EULAs are now commonplace features of children’s websites, they are rarely given the high visibility and degree of attention that they deserve. To ensure that these important legal contracts are not overlooked by users, they should be more prominently displayed on the site or more fully integrated into the primary content of the site (for example, a game could begin by taking users through the terms and conditions and privacy clauses in a fun but informative way). Furthermore, every effort should be made to facilitate children’s understanding of the various legal aspects of the contracts they are asked to enter into. While children’s-site designers are extremely talented at creating games and spaces that children are drawn to and deeply connect with, EULAs and privacy policies are not articulated in a language that is accessible to young audiences. A more child-friendly approach, such as that adopted by the UN in their articulation of the Convention on the Rights of the Child, would include both a child-friendly and adult version of the EULA to ensure that both children and their parents are able to access and understand the exact nature and scope of the terms and conditions they are expected to agree (and submit) to.

In establishing guidelines for an online community, policies are often in place for such practices as codes of conduct, membership requirements and communications style (Preece, 2000). As we have seen in the case studies examined above, policies (particularly EULAs) now also include clauses for copyright protection, as well as intellectual property claims over the entirety of the site’s contents (including those contributed by the user), aimed at generating profit and brand management, rather than fostering community interests. As online communities become increasingly important arenas for social development and cultural participation, the opportunities they provide to both users and commercial operators will continue to expand. So too, however, will underlying tensions between human rights and current market-research practices expand. The economic models of
online communities and games must therefore address human rights issues—
through the inclusion of the privacy-protection frameworks offered by COPPA
and PIPEDA, for instance—if a conflict of interest between users and developers
is to be avoided.

The uses and potential abuses of personal information require, first, an
explicit and clear campaign of awareness regarding the uses to which data may be
put in the storage and mining process. This places the onus on those who seek to
use personal data. The principle of fairness demands the promotion of awareness.
Providing “data subjects” with information concerning how their personal data is
used grants them the opportunity to resist or negotiate the uses of their informa-
tion. With opportunity comes the recognition that despite processes of “deindivid-
ualization” (Vedder, 1999) there remains embodied personhood. Though one of
the consequences of data mining is categorical profiling, the revival of embodied
personhood is another key consequence. Although social sorting and categoriza-
tion have the effect of dividing or separating the embodied individual from data
images, those data images have effects, firstly on life changes and secondly, on the
embodied individual they were abstracted from. In the case of children, it is espe-
cially vital that adequate principles are formulated and that appropriate guidelines
are established for ensuring that the child’s consent (as well as the parent’s con-
sent, if the subject is under the age of fourteen years) is not only obtained, but also
that it consists of truly informed consent. In the U.S., for example, the Code of
Federal Regulations has determined that “Without certain knowledge about age
and competence . . . researchers cannot be sure if subjects truly understand the
risks and procedures to which they are explicitly or tacitly agreeing to accept,” and
therefore children are unable to give informed consent to participate in Internet
research (Walther, 2002, p. 212). While current measures aimed at securing the
consent of children and their parents are inadequate and inconsistent, data miners
and market researchers may yet draw upon the standards and ethical guidelines—
previously established by academic institutions, governments, and trade
associations—that currently regulate research conducted on children. The Asso-
ciation of Internet Researchers (AoIR), for instance, proposes that “the greater the
vulnerability of the [subject] . . . the greater the obligation of the researcher to
protect the [subject]” and provides a framework for obtaining consent that includes
highly detailed consent forms and information sheets to be read and signed by both
the child and a parent (Ess & AoIR Ethics Working Committee, 2002).

As a result of the development of online game sites such as Neopets.com,
young people from all corners of the world can participate in a shared, interactive
community. Accessing such a site carries its costs, however. In order to participate
and gain access to the tools and play spaces offered by companies like Neopets
Inc., young users must first agree to relinquish large amounts of their personal
information. Although website privacy policies are in place on the various sites
analyzed, the flow of this data is nonetheless questionable. For while COPPA,
PIPEDA, and other attempts to regulate child-oriented websites consist of impor-
tant first steps toward the protection of children’s rights within the online environ-
ment, their effectiveness is already greatly hampered by technological advances in data mining capabilities, as well as by the innovative and highly adaptive strategies of market researchers. Given the vast and technical complexities involved, ongoing research in this area is needed to further explore the issues raised in this analysis, including the implications of data mining, privacy, organizational control, and the division between the rhetoric and reality of how information-gathering and information-processing practices are carried out on a daily basis. Further study of the activities and interactions children engage in while participating in these sites is also needed to more fully understand the nature of the growing relationship between children and data miners, and to determine how existing and future policies may address these emerging issues. As international conventions and national law have previously established that minors warrant special status and consideration in such commercial (as well as legal) matters, the lack of attention that has thus far been granted to the exchange of information occurring between children and marketers must be remedied if children’s rights are to be effectively upheld and protected online (Grimes, 2005). If children’s online contributions and activities are to continue to be used as fodder for market researchers, Internet and media policies, as well as accepted ethical standards for online market research conducted on children, need to be updated and revised. If children’s cultural participation is something we truly wish to foster and promote, then the appropriateness and potential consequences of these exchanges need to be re-evaluated and redefined through rigorous public and legal debate.

Acknowledgments
The authors would like to acknowledge the inclusion of a condensed version of this paper in the proceedings of the 38th Hawaii International Conference on System Sciences (IEEE), held January 3-6, 2005, Big Island, Hawaii (Chung & Grimes, 2005).

Notes
1. “Peer-to-peer” or “viral” marketing refers to the practice of using peer pressure and tendencies toward conforming to one’s peer groups to spread marketing and advertising messages. The process involves identifying opinion leaders in a particular group and enlisting them as “brand champions,” either through direct employment or by offering rewards and other incentives, using their status within the group to advertise the product or brand. The tactic relies on word of mouth and other informal communication channels to spread knowledge of the product or brand across the target market.
2. Not included in the present analysis is the Children’s Listbroker Privacy Act, new legislation first introduced in the U.S. in July 2004 as part of the Parents’ Bill of Rights (a set of nine legislative proposals aimed at granting parents more control over commercial influence on children). As this bill has not yet been passed, its relevance to the current discussion remains to be seen.

References


Jenkins, Henry. (2001, January-February). The kids are all right online. Technology Review; 104(1), 121.


**Websites**
- EverythingGirl.com
- Neopets.com
- Postopia.com
- OpinionSurveys.com