New Media & Society
http://nms.sagepub.com/

The archive and the processor: The internal logic of Web 2.0

New Media Society published online 13 May 2011
DOI: 10.1177/1461444811401735

The online version of this article can be found at:
http://nms.sagepub.com/content/early/2011/05/12/1461444811401735

Published by:
SAGE
http://www.sagepublications.com

Article

new media & society
1–17

DOI: 10.1177/1461444811401735
nms.sagepub.com

The archive and the processor: The internal logic of Web 2.0

Robert W. Gehl
The University of Utah, USA

Abstract
In Web 2.0, there is a social dichotomy at work based upon and reflecting the underlying Von Neumann Architecture of computers. In the hegemonic Web 2.0 business model, users are encouraged to process digital ephemera by sharing content, making connections, ranking cultural artifacts, and producing digital content, a mode of computing I call 'affective processing.' The Web 2.0 business model imagines users to be a potential superprocessor. In contrast, the memory possibilities of computers are typically commanded by Web 2.0 site owners. They seek to surveil every user action, store the resulting data, protect that data via intellectual property, and mine it for profit. Users are less likely to wield control over these archives. These archives are comprised of the products of affective processing; they are archives of affect, sites of decontextualized data which can be rearranged by the site owners to construct knowledge about Web 2.0 users.

Keywords
Amazon Mechanical Turk, archives, Digg, Facebook, labor, processing, storage, the Von Neumann Architecture, Web 2.0

During Mark Zuckerberg's first profile on CBS's *60 Minutes*, he helped reporter Leslie Stahl create her own Facebook profile (60 Minutes, 2008). He guided her through the template, even doing the work of typing in her 'likes' for her. 'Within a few minutes,' Stahl reports, somewhat surprised, 'I got a friend request' from someone she had not talked to in two years. Moments of inputting data into Facebook thus resulted in the

Corresponding author:
Robert W. Gehl, 2515 LNCO, 255 South Central Campus Drive, Salt Lake City, UT 84112-0491, USA
Email: Robert.gehl@utah.edu

elimination of years of lost time. Stahl notes that the near-instantaneous connection to friends is a reason why Facebook is so 'addictive.'

Speed, the new, and immediacy appear to be at the heart of Facebook, along with nearly every other Web 2.0 site. I define 'Web 2.0' as the new media capitalist technique of relying upon users to supply and rank online media content, then using the attention this content generates to present advertisements to audiences. It is currently the hegemonic business practice on the Web, employed by a wide range of sites, including Twitter, Google, YouTube, and Digg. As is evident from their interfaces, these sites are dedicated to immediacy, social connection, and instant access to information, much to the delight of users such as Stahl.

And yet, pushing past the glossy, AJAX-driven interfaces of Web 2.0, we confront another element of this business practice. Web 2.0 sites are not simply surfaces dedicated to immediacy; they are also comprised of vast server farms, rooms of computers humming away. Of course, these servers provide some of the processing power that drives the immediacy of a Web 2.0 site. But they also provide a function extremely necessary to any Web 2.0 business plan: rationalized storage of vast amounts of data. In other words, while Stahl constructs her profile, Zuckerberg's servers are busily storing her data. Here, we confront a contradiction: the smooth interfaces that users enjoy appear to be comprised solely of immediate connections and instant information, but the servers powering them are maintained in large part due to their long-term, archival potential. This contradiction is the motor that drives Web 2.0.

If we open those servers, we see that the Web 2.0 contradiction has its roots in the development of the modern computer itself, which is a synthesis of the immediate (in the form of the CPU or processor) and the archival (in the form of memory and storage of data). This fundamental architectural logic has informed the design of Web 2.0, not just its technical facts, but also its social structure. The fundamental architecture of the computer must, therefore, be linked to the logic of Web 2.0, because there is a social dichotomy at work based upon and reflecting (if not directly determined by) this architecture. In the hegemonic Web 2.0 business model, users are encouraged to focus on the new and the immediate. They are expected to process digital objects by sharing content, making connections, ranking cultural artifacts, and producing digital content, a mode of computing I call 'affective processing.' In essence, this business model imagines users to be a potential superprocessor. With enough users aggregated together via network effects and presented with a smooth interface (preferably something simple and binary such as a 'Like,' 'Tweet,' or 'Digg' button), they become a valuable source of digital artifact processing.

In contrast, the archival possibilities of computers are typically commanded by Web 2.0 site owners. They surveil every action of users, store the resulting data, protect it via artificial barriers such as intellectual property, and mine it for profit. This mode of new media capitalism prompts site designers to build websites that are capable of inscribing user activity into increasingly precise databases. Due to many sites' Terms of Service agreements, users cannot control these archives. These archives are comprised of the products of affective processing; they are archives of affect, sites of decontextualized data that can be rearranged by the site owners to construct particular forms of knowledge about Web 2.0 users.

The impact of this sociotechnical dichotomy is tremendous. If Derrida (1996), Foucault (1970, 1972), and Bowker (2005) are right in arguing that control of the archive leads to social power, then Web 2.0 site owners are becoming quite powerful because they have the ability to pull data from their archives to produce knowledge. New media capitalists seek to exchange these archives of affect with third parties (most commonly advertisers and marketers) to gain greater amounts of the classical storage unit of social power: monetary wealth. Thus, ultimately this paper argues that Web 2.0 sites are not simply spaces where users take control of content creation by constantly updating sites with new videos, Tweets, status updates, and Diggs; they are also devices designed to capture the affective labor of users and create archives of the digital material they produce.

To explore the Web 2.0 contradiction between the immediate and the archive, I first outline the roots of the processor/storage dichotomy in the Von Neumann Architecture approach to computer design. I then examine how new media sites encourage users to value the new and to engage in the affective processing of digital artifacts. I include three brief case studies: the NASA Clickworkers project, Digg, and the Amazon Mechanical Turk. Next, I explore how Web 2.0 sites archive the products of user-generated affective processing. I draw on Marx's Money-Commodity-Money' (M-C-M') circuit to illustrate how archiving user activities is a means to build social power. Finally, I conclude by examining the power of archives.

The Von Neumann Architecture

The basic architecture of computers involves the separation of the processor and storage. This architecture, dating to the mid-1940s and commonly called the 'Von Neumann Architecture,' calls for computer designers to store data and programs in a memory core, and to process that data and execute those programs with the processor (Eckert, 1945; von Neumann, 1993). It was first described by mathematician John von Neumann and used on the prototype EDVAC, the first stored-program computer[1] and the predecessor of all modern machines. In this logical organization, the storage unit of the machine and its processing unit are related to one another in a linear hierarchy of 'fetch-execute,' where the processor fetches data from storage, manipulates it, and then moves on to the next line of data. The processor only focuses on the immediate data it is working with, whereas the storage unit contains all the computer's command code and data.

In this architecture, the processor focuses on speed and discrete operations. It manipulates small chunks of data as quickly as possible, moving sequentially through each element of complex equations. The faster the processor moves through each instruction, the faster it produces results for users. The processor is thus a mechanical/electrical replacement for the collected labor of large groups of human computers, a prior form of information processing used since the 1700s (Grier, 2005). These groups of human computers were deployed in a division of 'mental labor' (Babbage, 1832) organized around mathematical operations; some would divide, some would multiply, and some would calculate square roots in a factory-inspired assembly line. Similarly, John von Neumann's description of the processor divides its functions along the logical lines of

mathematical operations in order to increase its speed (1993: 11–19). This internal division of labor is now a standard feature of processors (Aspray, 1990: 33).

Improvements in processor speed have altered user perceptions over the history of computing. By the 1960s, advanced computer designers strove to make the computer feel as if it were reacting immediately to the whims of the user, a mode of computing called 'real time' processing. Paul Ceruzzi (2003) argues that this 'mental model' of immediate computing influenced our contemporary computing culture, especially in the design of end-user software such as operating systems. Computer operating systems are now often judged on how well they utilize the full power of the processor and how quickly they complete multiple tasks.

In contrast, memory is an element of the archival potential of the computer. Its development is based on increasingly shifting data out of time. In the 1940s, while engineers strove to have the machine process data as fast as possible, in many cases (specifically those times that a remainder had to be carried over in a mathematical operation) some data had to be delayed momentarily. The memory developed by J. Presper Eckert and used in the EDVAC was based on mercury-line delays. Like their name implies, these devices used the differences in the speed of sound traveling through different media to delay certain bits of information, transducing information through mercury and thus effectively storing it (Aspray & Campbell-kelly, 1997: 92–93). In the Von Neumann Architecture, this technique is used extensively to include not only short-term storage of numbers for operations but also long-term storage of computer programs. Computer memory thus can be conceived of as this time delay writ large. Data is taken out of time and stored as indefinitely as the medium will allow: a few seconds for the 1940s-era mercury delay line, to years, perhaps centuries, with disk drives and solid state drives.

Like the processor, developments in computer memory shifted the ways in which users interpreted the machine. During the 1960s, the days of the mainframe, data was most often stored external to the machine on punch cards. This data was toted to the machine, loaded, and then after it was processed, the machine produced calculations. The development of tape reels and core memory marked a transition from batch processing to Random Access Memory (RAM), a more efficient form of storage. With the advent of spinning disks and later solid-state drives, mass storage and access to data and instructions was possible. This feature was quickly adopted because it made computers much easier to modify for different tasks, increased the amount of data the processor could work on, and allowed storage of documents and digital artifacts.

Thus, often when we talk about a computer, we discuss these two contrasting facets: how fast can it process? How much data can it store? These are the basic architectural facts of the technology, the result of design decisions made over a half century ago. The computer, post von Neumann, is therefore a unique synthesis of immediacy and archival capacity.

This synthesis has been replicated on the internet. The dual logic of the processor and the archive animates and in part determines the current business practice and social structures of the Web. The challenge for Web media companies is to always have new content to gain relevance in search engines and attract viewers. New media capital is meeting this challenge with the business practice of Web 2.0. According to technologist Tim O'Reilly (2005, 2007), Web 2.0 is the practice of getting users to add value to a

website by having them build its content, thus accelerating the cycle of media production so that sites become dynamic, constantly updated sources of new material. Users of all abilities – from professional to semiprofessional to amateur – are asked to create videos, write blogs, post comments, and rank media objects. Web 2.0 sites such as Facebook, MySpace, Twitter, Google, Amazon, and Digg have enabled this constant production of content by ceding control over the immediate to users. They have essentially built empty templates and invited users to fill them in. Due to this practice, users now have unprecedented power over popular trends on the Web. However, the catch here is that Web 2.0 site owners have not ceded the other half of the computational equation: the archive. While users are defining trends and shaping the now, Web 2.0 site owners are carefully shifting user-generated content out of time; thus, site owners command the past, a past largely imagined to be an increasingly granular map of user desires. The architecture of Web 2.0 is not only comprised of empty templates; it also uses massive server farms to store the content and associated data that users produce.

Interfaces of the new: 'What are you doing right now?'

The first element, the emphasis on the immediate, is built into the interfaces of popular Web 2.0 sites. In Facebook, users are confronted with a prompt: 'What's on your mind?' Twitter asks users 'What's happening?' and MySpace asks 'What are you doing right now?' These prompts ask the user of these social media sites to react, to present his or her current 'status': I'm happy, I'm going to the airport, I'm listening to Radiohead.

This emphasis on the immediate is not limited to social networks, but is also seen in media sharing sites. Flickr's homepage presents visitors with a count of photos uploaded 'in the last minute.' YouTube's homepage features 'Videos being watched right now.' Vimeo has a videos being shown 'Right Now' tab on its main page. Hulu has a 'Recently uploaded' page, featuring the latest video uploads.

Blogs and comment fields are also sites of immediacy. They typically read in reverse-chronological order; the newest post is on top, with older posts pushed down the page. Likewise, comments fields on newspaper sites are often organized in reverse-chronological order. The new is always privileged and on top. To follow the threads of discussion, one must click through pages of comments and attempt to reconstruct a conversation back through time.

Of course, these sites are augmented by the developments of mobile computing and smartphones; these allow users to update their statuses, comment on digital artifacts, and upload content from wherever they can get onto their networks. Telcom company Sprint offers 'The Now Network.' Verizon asks, 'Can you hear me now?' Users seek out those networks that can keep them connected wherever they are so they can continue to engage with new information streams. This emphasis on speed is so compelling that mobile companies and software entrepreneurs are increasingly using location-aware software to immediately alert users to consumer opportunities as they navigate public spaces.

In sum, as Chris Gerben (2009) notes, Web 2.0's interfaces heavily emphasize the new *even at the cost of other modes of organization such as relevance or importance.* Similarly, David Berry (2008: 367) argues that network theory – a mode of inquiry often deployed by architects of Web 2.0 – 'privilege[s] a reading of reality that highlights the

synchronic dispersal over the diachronic unfolding,' and that 'Networks, in a certain sense, abolish history and shift our focus to the event, the happening or the now.' In their examination of MySpace, Coté and Pybus (2007: 101) argue that users of Web 2.0 sites are engaged in a 'never-ending process of becoming ... Each new device and resource expands the capacity of their 'digital body' and allows them to forge new compositions of relations.' This dual reliance upon user-generated 'newness' and the emphasis on always-becoming are built into the architecture of Web 2.0. It imagines subjects that are always connected, always updating, always searching, and never stopping their restless motion from one site to the next.

However, this emphasis on the new is not, in fact, new. Rather, it is latest in the longstanding sociotechnological development of computer processing. As Adrian Mackenzie (1997: 60) argued in the 1990s, the focus on the new was part of the two dialectical processes of the internet: the emphasis on 'real-time drives' and the archival impulse. He writes that 'The virtual ... can be positioned at the interactive threshold between the processes of real-time and the processes of the archive.' He correctly sees this dialectic in the Von Neumann Architecture of processor and memory. This dichotomy was built into the internet from its earliest days; as Andrew Flanagin et al. (2010) note: 'The defining characteristic of an end-to-end system [such as the internet] is that network "intelligence" (discrimination and processing functions) exists primarily at the periphery of the network, while the network pathways remain neutral, handling all data traffic identically.' The emphasis on real-time is thus a product of the internet's long-standing architecture that assumes an end-user who is interested in getting data fast. As Virilio (1995: 141, original emphasis) argues, '*the reality of information is entirely contained in its speed of dissemination ... speed is information itself!*' This emphasis is also based on the short-term goals of processing and the increasing speed of traffic on the internet, while the emphasis on the archive is part of a longer historical process, one that I will discuss below.

With the advent of broadband technology, the network is achieving even greater speeds than it saw in the 1990s. Arising directly out of faster internet connections and new suites of Web programming technology such as AJAX (Asynchronous Javascript And XML), one of Web 2.0's most salient features is that it is as responsive as desktop software. In fact, in the marketing literature for Web 2.0 services, online software is presented as a *replacement* for desktop software. AJAX is a codification of a new relationship between server and client computer, where only the most immediately needed data is served to the client. In this environment, as AJAX manual writer Holdener (2008) puts it, 'The user will perceive everything about the web application as being self-contained. With this technology a savvy developer can make an application function in virtually the same way, whether on the Web or on the desktop.' A well-designed website utilizing AJAX only requests from the server the information the user is currently interested in; the entire site does not have to reload. Thus, Web 2.0 site designers seek to replicate and surpass the surface-level immediacy of the desktop operating system (Governor et al., 2009: 127).

Web users are engaging with this immediacy by feeding updates into it and relying upon it to provide emotional contact instantaneously. As Sherry Turkle (2007) argues: 'We live a contradiction: Insisting that our world is increasingly complex, we nevertheless

have created a communications culture that has decreased the time available for us to sit and think, uninterrupted. We are primed to receive a quick message to which we are expected to give a rapid response.' That is, the speed at which our electronic networks can connect us to others creates a new relationship to emotion: 'Emotional life can move from "I have a feeling, I want to call a friend," to "I want to feel something, I need to make a call."' The emphasis on the new in Web 2.0 leads to immediate affective exchanges; I email you, you chat with me. If you do not, I become anxious. Why aren't you emailing me back *right now*? If Clay Shirky (2008) is right in arguing that our mediascape is marked by 'filter failure,' an environment where we are unable to filter all the possible content we might encounter, then perhaps this emphasis on the new is logical. A connection (however weak) with a friend *right now* might outweigh the value of terabytes of uncontextualized, fragmented data that may or may not offer emotional value but nevertheless demands that we sift through it.

However, this is not just a structure determined by the technological architecture of the computer, or by the actions and desires of users; it is also determined and extended by the needs of late capitalism. Agger (2004) has aptly named this formation 'fast capitalism,' arguing that it has radically extended Taylor's vision of scientific management. Virilio's (1986, 1995, 2005) arguments about dromology orient us to the use of fast-capitalist tools such as instant communications and instantaneous navigation of digital spaces. Digital environments condition users to expect information immediately and thus to act upon it. In some senses, this is a social good; it enables us to access greater spheres of information than our ancestors.

And yet, this phenomenon has not arisen in a social vacuum. When we consider this focus upon the new as another instance of the just-in-time demand for labor that marks late capitalism, particularly (but not limited to) affective immaterial labor, then this emphasis upon the new is clearly a case of Web media corporations relying upon users to do the work of quickly and cheaply processing digital artifacts to generate an informational and affective surplus. Users are relied upon contingently and intermittently, but relied upon nonetheless. While users have become accustomed to instantaneous action from their networked devices and instantaneous connections to their friends, capitalists, investors, and media companies have become accustomed to the near-instantaneous processing of data by users and have positioned themselves to exploit the results of this processing by building archives from user activities.

Crowdsourcing: from Mars to Digg to the Mechanical Turk

To illustrate this, I offer three examples that loom large in the mythology of Web 2.0. The first is a non-profit volunteer effort. In 2000, NASA began its Clickworkers project. This was a part-time project that allowed public volunteers to mark craters on photographs of Mars. Marking craters is a tedious and time-consuming task for an individual. According to Szpir (2002), 'The task is usually undertaken by someone trained in the art and science of rating craters, but there are many thousands of craters on the planet and, well, most scientists (even graduate students) have better things to do.' Seeking a more efficient way, the Clickworkers project was an experiment to see if the public volunteers could process those images online as reliably and faster than the scientists who would have

done the work. It was a resounding success. According to Benkler (2006: 69), more than 85,000 volunteers visited the site and made over 1.9 million entries. 'An analysis of the quality of markings showed "that the automatically computed consensus of a large number of clickworkers is virtually indistinguishable from the inputs of a geologist with years of experience in identifying Mars craters."' These contributions were done by part-time volunteers, many of whom spent five minutes on the site before moving on. As a part-time experiment, the project was staffed by one engineer with two consulting scientists, creating tremendous savings in time and resources for NASA. But more importantly, it demonstrated that the internet provides a structure for massively distributed human processing; users from all over the world lent a few minutes of their visual acuity to the project, and these micro-moments of labor and attention aggregated into an incredible superprocessor.

The Clickworkers project has been mythologized in the arguments of Web 2.0 enthusiasts like Benkler (2002, 2006; Benkler & Nissenbaum, 2006) and Howe (2006, 2008) (who coined the term 'crowdsourcing'). For Benkler and Howe, this development means that the production of knowledge has finally been 'democratized,' broken out of the confines of space, expertise, and certification. The Clickworkers project proved that users will volunteer to help an institution achieve a goal. Users' online activities are presented as 'spare computing cycles' (Howe, 2006), likened to the spare processing cycles of an idling processor. For Benkler and Howe, this also means that corporations must take advantage of users' free labor, since to choose otherwise would be to make an irrational business decision. Distributed, networked labor, they argue, is now much cheaper for capital to rely upon. This iteration of capitalism, dubbed by Benkler (2006: 3) the 'networked information economy,' involves 'decentralized individual action – specifically, new and important cooperative and coordinate action carried out through radically distributed, nonmarket mechanisms that do not depend on proprietary strategies.' While this is presented as 'nonmarket,' it is clear that Benkler and Howe see this development as, in fact, a new, cheap labor market.

Corporations have responded to this market. Social networks, video sharing sites, auction sites, and search engines rely upon the labor of users to create their content. Here, I want to focus on two examples of for-profit distributed human computing.

Digg

As discussed above, Shirky (2008) has argued that users of the Web (and other media) suffer from 'filter failure'; consumers are drowning in media objects and are unable to discern which are relevant. Digg is a response to filter failure. The service offers users a way to sift through the mass of digital material on the Web. This is accomplished by the work of users who do one or more of three tasks: submit material, rate it (a process called 'Digging' or 'Burying'), and comment upon it. If an item gets enough positive 'Diggs,' it reaches the front page, where millions of visitors can see it, link to it, and comment upon it. Conversely, items can get 'buried' by Digg users, either because they are irrelevant, not entertaining, or spam. In addition, the submissions are also categorized by users into subsections such as Technology, World and Business, and Gaming, each with their own subsections. Thus the vast material available on the Web can be sorted and rated, presenting a structured snapshot of what is popular online.

Digg co-founder Kevin Rose (MacManus, 2006) argues that this system returns power to 'the masses': 'This was the first time that anyone experimented with allowing the general mass audience to decide what they believed to be the most important topic of the day.' Indeed, in many descriptions of the site, it is as if there are no administrators, investors, or site owners at all; as *How Stuff Works* writer Layton (2006) presents it, the only agents involved in the site are varying grades of users, from casual to 'dedicated.' Even her description of the server–client structure of the site – a complex arrangement of hardware and software that requires IT labor to run it – elides any other persons laboring on the site. Users are ostensibly the only ones in control.

The results of this user-led sorting, ranking, and surveillance are distributed across the Web in widgets that proclaim that the news items they contain are 'Powered by Digg's Users.' Users are thus explicitly compared to a microprocessor, implying that the millions of Diggers who sort and rank items are more powerful than any hardware-based processor or software algorithm. These widgets offer an easily understood numerical assessment of its highly rated items; each has a Digg count. Moreover, they offer speed and the new: as Digg co-founder Jay Adelson (Business Week, 2005) puts it:

> [Digg] attracted the attention of the news media immediately – the fact that we had this incredible speed. Automated systems take time to crawl the net. Editorial systems have the human factor. They may decide they're not interested that day, or they'll do it tomorrow. In our case, there's no barrier, so the second a story would be interesting to this mass public, we can break it.

This emphasis on the new in news appeals to those Web users who seek immediate access to information. Without such a filter, this argument goes, users might miss out on breaking news stories because they are navigating serendipitous content in sources such as newspapers.

Thus, Digg is built upon the model that the NASA Clickworkers project pioneered: distributed processing comprised of users clicking their way through news stories. Like the Clickworkers project, Diggers need not spend more than a brief moment 'digging' a story; the aggregation of these micro-moments of labor produces the Digg home page.

Amazon Mechanical Turk

Digg's deal with users is based on the exchange of micro-moments of labor for the wage of prefiltered content, not unlike the exchange between broadcasters and audiences that Dallas Smythe (1981) detailed. Thus it offers no financial renumeration for labor. A site which takes this model and injects monetary wages is Amazon's Mechanical Turk. The Mechanical Turk is a marketplace of tasks, which Amazon calls 'Human Intelligence Tasks' or HITs. As the name implies, HITs emphasize those tasks, such as image recognition or audio transcription, that require human judgment. In essence, 'Turkers' who complete HITs are marketed to employers as the world's best computer, combining the unparalleled capacity of humans who can read, recognize images, and make immediate and accurate judgments. As Barr and Cabrera (2006) explain, Amazon envisioned the service as an answer to companies that need meta-data improvement,

image selection, and translation to be done on increasingly large scales. Computers cannot handle these types of tasks with any accuracy, and hiring workers to do them in-house is extremely expensive. Thus, the Mechanical Turk has utilized the Web as a means to connect companies in need of repetitive digital tasks to a worldwide labor market of micro-laborers.

Despite its reliance on human-processors, the Amazon Mechanical Turk's marketing literature abstracts the *human* processing that takes place during HITs. Amazon wryly calls this 'artificial artificial intelligence,' referencing the interface, which makes human work look mechanical. It is structured not unlike the server–client practice of networked computing: the employer sends a request to Amazon, and the humans' response to the request is served back via AJAX-style programming. The legendary marketplace, where labor meets capital in a personified negotiation, is replaced by a screen interface, where labor finally becomes completely mechanical and rationalized. Human labor is reduced to cost, a mere input in the production process, and a cheap one at that. In many cases HITs are worth a few cents (US) a task. For example, as of this writing, one HIT asks Turkers to classify advertisements for $0.05 in three minutes. Another asks Turkers to 'check if these websites work' for one penny a piece.

In sum, like Digg, the Mechanical Turk is built upon the Clickworker model, but takes that model further by emphasizing the processing of digital artifacts and de-emphasizing knowledge of what these tasks are for, much like the divorce between conception and execution Harry Braverman (1975) describes. In this way, Turkers are encouraged to ignore everything but the micro-labor task at hand. While humans-as-laborers are elided in the structure of Mechanical Turk, employers are also hidden behind layers of abstraction. As Jonathan Zittrain (2009) explains, Turkers do not have much knowledge of their employers. They simply have Amazon accounts and receive micro-payments for services rendered. From the employers' perspectives, users are imagined as processors, meant to do tasks quickly and accurately and return the results – without question – to unseen entities.

While the Web 2.0 emphasis on 'the wisdom of crowds' is compelling, the goal of commercial Web 2.0 sites is to capture the processing power of a critical mass of users, either directly (as in the case of Digg) or indirectly (as in the case of Amazon Mechanical Turk). Often, this processing is evacuated of ethics; the owners of the sites do not particularly care *what* the users are processing, so long as their attention is fixed upon the site. In short, the development of Web 2.0 is a trajectory of increasing capitalization of the processing power of the masses of computer users. Whereas computer engineers might have dreamed of building truly universal machines, ones that could fully replace humans, computers still do not compete with a mass of humans. No computer can compete with us when we join together and tackle problems. The question is, what do we do with this capacity? In Web 2.0, what began as an ethic of nonprofit volunteering to a greater cause (NASA Clickworkers) has been morphed to an individualistic emphasis on sharing and personal connection (Digg, Facebook, MySpace, Twitter, YouTube) and even to the ultimate just-in-time flexible labor market (Amazon Mechanical Turk). This emphasis is reinforced by the predominant focus on the new. The user has to update her status, check on her friends, make new friends, recheck for a new connection or emotion, while 'Turkers' seek the latest HIT. In this milieu, users are imagined to be the processors

computers never could be. However, computers do have humans trumped in another area: memory.

Building an archive of culture and affect

While computer scientists could not replace human skills such as image recognition and subjective rankings with artificial intelligence, the other half of the computer's architecture has been much easier to construct, expand, and improve upon. Memory is as essential to modern, von Neumann-inspired computers as is the processor. The processor works on data, but data (in the form of instructions and results) must be stored somewhere. Memory capacity has grown tremendously, leading to today's terabyte drives that store vast amounts of information. This information must be routed to the processor. To do so, computer architects have developed busses, short-term caches of memory, and dedicated pathways for instructions and data in order to link them. Thus, we have a basic architecture: processor, memory, and the path between the two. Computer engineers seek to optimize the relationship between memory and the processor to create an ideal synthesis of the immediate and the archival.

In Web 2.0, the path between the user/processor and the archive is the broadband internet connection. Broadband connections are the sine qua non of Web 2.0. Without them, AJAX-based applications that replicate desktop software would not be viable. With them, these applications work and users can readily access them. Whereas dial-up connections had to be established by dialing a number and connecting, a process that can be unreliable and at the very least ties up phone lines in many people's homes, broadband connections such as DSL, cable, and WiMax can always be on. The connection becomes silent (i.e., no more chirping sounds over phone lines) and invisible, since it does not get in the way of the user's online experience. Moreover, since this constant connection is far more reliable than dial-up, it is akin to the dedicated busses installed between memory and processors within the Von Neumann Architecture. With this bus, site creators can imagine masses of users who will interact with digital material without worrying about the connection. Thus, sites such as Digg and the Mechanical Turk can rely on users who are focused only upon completing micro-tasks.

In addition, broadband not only enables the distributed human processing that the Clickworkers project, Digg, or Amazon requires, it also enables the *storage* of the results of human processing. The data that users process must be stored somewhere. This is an often overlooked aspect of always-on broadband connections. While broadband is very often presented to consumers as a fast way to *download* material, every download also requires *uploads*. At the very least, a client computer must upload a request, such as the XMLHttpRequest object, to a server to receive data. Thus even a user who 'free rides' on a site, only browsing but not contributing content, generates such data (Flanagin et al., 2010: 186). These requests can easily be stored by the server, forming an archive of user activities that can be later analyzed and data-mined. Moreover, Web users rely on broadband connections to upload photos, movies, or blog posts. Just as data is necessarily and automatically migrated from memory to the processor and back in the Von Neumann Architecture, storing the results of user activities in Web 2.0 is a built-in process. Capturing user activities in matrices of server-side request logs, XML meta-data,

and IP address logs is a *necessary* aspect of the broadband/AJAX connection between client and server. As users surf the interfaces of Web 2.0, the online archive grows ever more precise.

A-P-A'

Ultimately, for the owners of social media sites, the goal is to store as much user-generated content and data as possible, serve it to users who process it further, and then store the results, creating an ever more precise and extensive archive. Facebook is a prime example of this. In order to grow, it requires more participants to attract other participants. This is the so-called 'network effect,' where a networked technology's value grows as more people use it. The network effect is apparent in any communications network; telephones, for example, are only useful if there's someone to call. Digg, Facebook, YouTube, and other social media take this a step further by expanding possible uses. On these sites, users do not simply email one another, but play games, chat, give gifts, comment, post media objects, and display their statuses. These interactions are often (but not always) asynchronous. Users often interact not with one another in real-time but rather with digital ephemera that stand in for users: avatars, status updates, images, and videos. Thus, what social media site users are interacting with is an archive of affect, digital objects that have meaning within the context of social connections. They are processing this digital archive: sorting their contacts into lists, liking this status update, commenting on that photograph, or sharing a virtual gift.

Facebook seeks to have a large archive (A) of these objects for users to interact with. Facebook was initially seeded with applications such as the Wall (an area for user comments), photo sharing, and notes. These basic applications allowed users to post text, photos, and comments on other users' profiles. As users interact with these objects, processing (P) them, Facebook watches their actions and collects data, archiving (A') this newly generated data. This is the information Facebook seeks to sell to advertisers. The process has been accelerated as Facebook has opened its Application Programming Interface (API) to third-party developers who create more applications inside which users interact. In sum, Facebook – and other Web 2.0 sites – seek to grow the archive through the process A-P-A'. The larger the archive, and the more granular the data about the desires, habits, and needs of users, the more valuable the archive. And if the archive is reliably linked to users who can sort data and process digital artifacts, the archive can be grown and made more precise.

Each of these steps is highly necessary, but only one can cause the archive to grow. As in the Marxian Money-Commodity-Money' formula that this formula echoes, the process that grows the archive is labor, in this case the micro-labor of users whom Ursula Huws would call 'cybertariats' (2003), since the work in question is often highly casualized and even presented as entertainment. This is part of the larger exploitation of previously untapped 'people power' on the Web, where the leisure of Web users who seek entertainment and diversion is finally made productive for globalized capitalism (Fisher, 2010: 137–143). Whether they are Digging, Turking, or simply updating their statuses, users are explicitly imagined to be the labor/processor core that 'powers' social media. They are the 'Intel Inside' of Web 2.0. In sum, they are the processor in the Von

Neumann Architecture, a social reflection of the internal division of labor that constitutes computers.

Conclusion: Archives and power

One of the major tropes of Web 2.0 is that websites organized with users making decisions eliminates editorial and mass media authority, thus democratizing media production (P. Anderson, 2007; Beer and Burrows, 2010; Benkler, 2006; Bruns, 2008a; Richards, 2009). Web 2.0, this argument goes, raises the average user to the level of editor and central authority. It removes gatekeepers, allowing average users to produce, evaluate, and distribute content. The future, as Axel Bruns (2008b) proclaims, is 'user-led,' no longer the domain of executives who plan broadcast schedules and distribute media from centralized studios. We are no longer beholden to the tyranny of mass media, argues Chris Anderson (2006); now we can find whatever entertainment we desire in affective niches located somewhere on the 'long tail' of participatory content creation.

However, authority is not eliminated in this new media environment. While Web 2.0 may have in fact created new ways for users to find and manipulate digital content, the archival capacity of Web 2.0 allows for new centralizations of power, hidden away beneath the abstractions of the smooth Web 2.0 interface. Although traditional mass media gatekeeping roles may have been eroded, Web 2.0 has enabled new media companies and entrepreneurs to assume a curatorial role (Gehl, 2009); these curators build archives out of the products and traces of users' affective processing, protect them via Terms of Service agreements and intellectual property regimes, and mine them for profit.

For example, the Facebook TOS states:

> You own all of the content and information you post on Facebook, and you can control how it is shared through your privacy and application settings. For content that is covered by intellectual property rights ... you specifically give us the following permission, subject to your privacy and application settings: you grant us a non-exclusive, transferable, sub-licensable, royalty-free, worldwide license to use any IP content that you post on or in connection with Facebook ('IP License').This IP License ends when you delete your IP content or your account *unless your content has been shared with others, and they have not deleted it*.[2]

All is equitable until the last clause of the last line. Facebook claims no ownership over user intellectual property (assuming the user sets the privacy controls correctly). Facebook even will relinquish any claims to their licensed use of user material after account closure. *Unless it has been shared*. Since the explicit purpose of Facebook is to allow users to share their photos and writings, Facebook cleverly captures user data in a perpetual license while denying its intention to do so. Unless the user's 'friends' also delete the shared data, it will always be licensed to Facebook. Facebook is a service allowing users to share among their 'social graphs,' but this is itself simultaneously an expression of a second, less explicit purpose of the site: you may share with others while we capture the digital objects you share in order to gather data on your preferences and desires.

But data sets are not in themselves archives. To be an archive, the material collected must be done in an organized manner that allows for the *post hoc* construction of power/

knowledge. 'Indeed, how could one start constructing an archive without knowing the principle of its construction, without knowing in advance, among other things, what to select?' (Chang, 2010: 204). The material collected must be done in anticipation of its future reconstruction. Briankle Chang (2010: 205) sees the archive as existing in the future perfect: 'they will have become what they already were.' This becoming is always-already waiting for the *archon* (authority, curator) to appear as predicted in the future perfect. As Bowker (2005: 18) argues, 'what is stored in the archive is not facts, but disaggregated classifications that can at will be reassembled to take the form of facts about the world.' Thus, what is required is an authority to construct 'facts' from the fragments that sit on the archive's shelves. Bowker's name for our computer-driven memory episteme is 'potential memory,' a mode of power where those with access to the archive create narratives post hoc from a priori taxonomically organized objects that are scattered across many physical storage sites.

Web 2.0 lends itself to such post hoc constructions. Marketers, lawyers, entrepreneurs, social scientists, psychologists, and experts in so-called 'big data' have built the Web 2.0 archives in order to construct exchangeable images of user/consumers. The 'facts' that will become produced in Web 2.0 are largely concerned with consumer preferences. Whereas state-based interpellation of identities might arise from the metrics of security (date of birth, race, country of origin), rationalized identities in Web 2.0 arise from the metrics of capital and consumption: user profiles, categorized social connections ('friends,' 'co-workers,' 'family'), credit scores, searches, purchase histories, media consumption, desires, fantasies, demographics, and movements through space (Andrejevic, 2007); i.e., this is Deleuze's (1992) 'dividuation' in action. As far as marketers and investors are concerned, these are the most salient digital fragments to be stored in the servers of these sites. However engaged users are with their Tweets, profiles, articles, videos, and images, in this adaptation of the von Neumann division of computational labor, users are often reduced to affective processors working for the owners of the digital archive.

Thus, although some popular and academic accounts of Web 2.0 often present this as a form of media that eliminates editorial authority, by considering Web 2.0 as an expression of the relationship between users/affective processors and the owners of digital archives, we can readily see that authority is alive and well online, transcending the internet and becoming neo-Hobbesian sovereigns that Jarod Lanier (2010) calls 'the lords of cloud computing' who command data flows and storage. Although editors and gatekeepers have seen their roles eroded, data-miners have emerged as the new personification of media power. As Vincent Mosco (2004) argues, in the history of media technology in capitalism, power always reasserts itself in some form, despite the utopian proclamations of democracy and equality that accompany a new media form.

Rather than laud Web 2.0 for its anti-authoritarian, disruptive properties, future critical work must emphasize the archival side of the Web 2.0/Von Neumann Architecture because archives are sources of political power. As Derrida (1996: 4, note 1) argues: '[T]here is no political power without control of the archive, if not memory. Effective democratization can always be measured by this essential criterion: the participation in and access to the archive, its constitution, and its interpretation.' If this is the case, Web 2.0 sites should be judged by the ways in which they allow democratic access to their

archives. Judging them by evaluating the ways in which they allow users to 'be the media' or resist mass media authority is not enough. With Derrida's criterion in mind, most Web 2.0 sites are totalitarian because their archives, as well as the conditions of production of social facts based upon those archives, remain closed to the very users that have built them.

Acknowledgments

This research was funded by a dissertation completion fellowship from the College of Humanities and Social Sciences, George Mason University. I benefited from critiques from Hugh Gusterson, Alison Landsberg, Tim Gibson, Mark Sample, and the two anonymous reviewers.

Notes

1. Technically speaking, Maurice Wilkes's EDSAC was the first operational stored-program computer, beginning operation two years prior to EDVAC. However, the plan for the EDVAC was the first time a practical stored-program computer was proposed, and parts of EDVAC were demonstrated to a small group of observers prior to the EDSAC.
2. Available at http://www.facebook.com/terms.php, last accessed 13 October 2010; my emphasis.

References

60 Minutes (2008) *The Face behind Facebook*. Available at: http://www.cbsnews.com/video/watch/?id=3706601n&tag=related;photovideo

Agger B (2004) *Speeding up Fast Capitalism: Internet Culture, Work, Families, Food, Bodies.* London: Paradigm Publishers.

Anderson C (2006) *The Long Tail: Why the Future of Business is Selling Less of More*. New York: Hyperion.

Anderson P (2007) What is Web 2.0? Ideas, technologies and implications for education. *JISC Technology and Standards Watch* (Feb): 1–64.

Andrejevic M (2007) *iSpy: Surveillance and Power in the Interactive Era.* Lawrence: University Press of Kansas.

Aspray W (1990) *John von Neumann and the Origins of Modern Computing.* Cambridge, MA: The MIT Press.

Aspray W and Campbell-Kelly M (1997) *Computer: A History of the Information Machine.* New York: Basic Books.

Babbage C (1832) *On the Economy of Machinery and Manufactures, of 1835.* New York: M Kelly.

Barr J and Cabrera LF (2006) AI gets a brain. *Queue* 4(4): 29.

Beer D and Burrows R (2010) Consumption, prosumption and participatory web cultures: An introduction. *Journal of Consumer Culture* 10(1): 3–12.

Benkler Y (2002) Coase's penguin, or, Linux and 'the nature of the firm.' *The Yale Law Journal* 112(3): 369–446.

Benkler Y (2006) *The Wealth of Networks: How Social Production Transforms Markets and Freedom.* New Haven, CN: Yale University Press.

Benkler Y and Nissenbaum H (2006) Commons-based peer production and virtue. *Journal of Political Philosophy* 14(4): 394.

Berry DM (2008) The poverty of networks. *Theory, Culture and Society* 25(7–8): 364–372.

Bowker GC (2005) *Memory Practices in the Sciences.* Cambridge, MA: MIT Press.

Braverman H (1975) *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century.* New York: Monthly Review Press.

Bruns A (2008a) *Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage.* New York: Peter Lang.
Bruns A (2008b) *The Future is User-led: The Path towards Widespread Produsage. Fibreculture* (11). Available at: http://journal.fibreculture.org/issue11/issue11_bruns.html (accessed 14 December 2010).
BusinessWeek (2005) *How Digg Uncovers the News.* Available at: http://www.businessweek.com/magazine/content/05_47/b3960426.htm
Ceruzzi PE (2003) *A History of Modern Computing.* Cambridge, MA: MIT Press.
Chang BG (2010) To the archive: A postal tale. *Communication and Critical/Cultural Studies* 7(2): 202.
Coté M and Pybus J (2007) Learning to immaterial labour 2.0: MySpace and social networks. *ephemera* 7(1): 88–106.
Deleuze G (1992) Postscript on the societies of control. *October* 59: 3–7.
Derrida J (1996) *Archive Fever: A Freudian Impression.* Chicago, IL: University of Chicago Press.
Eckert JP (1945) Disclosure of a Magnetic Calculating Machine. Unpublished typescript: 525–539.
Fisher E (2010) *Media and New Capitalism in the Digital Age: The Spirit of Networks.* New York: Palgrave Macmillan.
Flanagin AJ, Flanagin C and Flanagin J (2010) Technical code and the social construction of the internet. *New Media and Society* 12(2): 179–196.
Foucault M (1970) *The Order of Things: An Archaeology of the Human Sciences.* London: Tavistock Publications.
Foucault M (1972) *The Archaeology of Knowledge; and, The Discourse on Language.* New York: Pantheon Books.
Gehl R (2009) YouTube as archive: Who will curate this digital *Wunderkammer*? *International Journal of Cultural Studies* 12(1): 43–60.
Gerben C (2009) *The Future of Democracy, Economy, and Identity in 21st Century Texts. Media in Transition 6.* Available at: http://web.mit.edu/comm-forum/mit6/papers/Gerben.pdf (accessed 15 December 2010).
Governor J, Hinchcliffe D and Nickull D (2009) *Web 2.0 Architectures.* Sebastopol: O'Reilly.
Grier DA (2005) *When Computers Were Human.* Princeton, NJ: Princeton University Press.
Holdener AT (2008) *Ajax the Definitive Guide.* Farnham: O'Reilly.
Howe J (2006) The rise of crowdsourcing. *Wired Magazine* 14(6): 1–4.
Howe J (2008) *Crowdsourcing.* New York: Random House.
Huws U (2003) *The Making of a Cybertariat.* New York: Monthly Review.
Lanier J (2010) *You Are Not a Gadget: A Manifesto.* New York: Knopf.
Layton J (2006) How Digg Works. *How Stuff Works*. Available at: http://computer.howstuffworks.com/internet/social-networking/networks/digg.htm
Mackenzie A (1997) The mortality of the virtual: Real-time, archive and dead-time in information networks. *Convergence* 3(2).
MacManus R (2006) Interview with Digg founder Kevin Rose, Part 1. *ZDNet*. Available at: http://blogs.zdnet.com/web2explorer/index.php?p=108
Mosco V (2004) *The Digital Sublime: Myth, Power, and Cyberspace.* Cambridge, MA: MIT Press.
O'Reilly T (2005) Not 2.0? *O'Reilly Radar*. Available at: http://radar.oreilly.com/archives/2005/08/not-20.html
O'Reilly T (2007) What is Web 2.0? Design patterns and business models for the next generation of software. *Communications and Strategies* 65: 17–37.
Richards D (2009) A social software/Web 2.0 approach to collaborative knowledge engineering. *Information Sciences* 179(15): 2515–2523.

Shirky C (2008) It's not information overload. It's filter failure. Available at: http://web2expo.blip.tv/file/1277460

Smythe DW (1981) *Dependency Road: Communications, Capitalism, Consciousness, and Canada.* Norwood: Ablex Pub. Corp.

Szpir M (2002) Clickworkers on Mars. *American Scientist*. Available at: http://www.americanscientist.org/issues/pub/clickworkers-on-mars

Turkle S (2007) *Can You Hear Me Now? Forbes*. Available at: http://www.forbes.com/free_forbes/2007/0507/176.html

Virilio P (1986) *Speed and Politics: An Essay on Dromology.* Trans. Mark Polizzotti. New York: Semiotext(e).

Virilio P (1995) *The Art of the Motor.* Minneapolis: University of Minnesota Press.

Virilio P (2005) *Negative Horizon.* New York: Continuum.

Von Neumann J (1993) First Draft of a Report on the EDVAC. *IEEE Annals of the History of Computing* 15(4): 27–43.

Zittrain J (2009) Minds for sale. *Berkman Center for Internet and Society*. Available at: http://cyber.law.harvard.edu/interactive/events/2009/11/berkwest

Robert W. Gehl is an assistant professor of new media in the Department of Communication at the University of Utah. He received a PhD in Cultural Studies from George Mason University in 2010.