Measuring the New Economy: Industrial Classification and Open Source Software Production

Fernando Elichirigoity* and Cheryl Knott Malone**

* University of Illinois at Urbana-Champaign, Graduate School of Library and Information Science, 501 E. Daniel Street, Champaign, IL USA 61820-6211, elichi@uiuc.edu

**University of Arizona, School of Information Resources and Library Science, 1515 E. First Street, Tucson, AZ USA 85719, ckmalone@u.arizona.edu

Fernando Elichirigoity is an Assistant Professor in the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. His teaching includes Information and Globalization, Information Consulting, Wireless Technology and Society, Competitive Intelligence, and Business Reference. His areas of research include social construction of classification systems, globalization and information technologies, technologies of attention and persuasion.

Cheryl Knott Malone is an Associate Professor in the School of Information Resources and Library Science at the University of Arizona where she teaches courses on the organization of information, government information, and information literacy instruction, among others. Her research interests include the social, political, and cultural construction of classification systems and the history of libraries and print culture.

Elichirigoity, Fernando and Cheryl Knott Malone. **Measuring the New Economy: Industrial Classification and Open Source Software Production.** *Knowledge Organization, 32(3).* 117-127. 32 refs.

ABSTRACT: We analyze the way in which the North American Industry Classification System (NAICS) handles the categorization of open source software production, foregrounding theoretical and political aspects of knowledge organization. NAICS is the industry classification scheme used by the governments of Canada, Mexico and the United States to carry out their respective economic censuses. NAICS is considered a rational system that uses the underlying economic principle of similar production processes as the basis for its classes. For the Information Sector of the economy, as formulated in NAICS, a key production process is the acquisition and defense of copyright. With open source, copyleft licensing eliminates copyright acquisition and protection as major production processes, suggesting that the open source software industry warrants a separate NAICS category. More importantly, our analysis suggests that NAICS cannot be understood as a taxonomy of objective economic activity but is instead a politically and historically contingent system of data classification.

1. Introduction

This paper is a critical analysis of the constructed nature of the classes constituting the North American Industry Classification System (NAICS), a system originally created to measure the shift from manufacturing to service industries, a shift often called "The New Economy." This work extends and deepens our previous explorations of NAICS. In Malone & Elichirigoity (2003) we argued that underlying the NAICS depiction of the "Information" sector of the economy is a definition of a knowledge or information product as one that is subject to copyright. In this paper we argue that NAICS actually misses im-

portant aspects of the New Economy by limiting "Information" in such a way. We extend the analysis by looking at how NAICS deals with the creation of significant value in the production of open source software, such as the Linux operating system, which is not necessarily measurable in pecuniary terms, and which can function as political contestation as well as economic activity. In fact, the open source movement can be seen as both a deliberate effort to counter the increasing protection of copyrighted information at the expense of use, even fair use, and as an emergent mode of production. By foregrounding the nature of open source production we bring attention to forms of economic activity that problematize assumptions of neoliberal economic theory embedded in NAICS. We illustrate these arguments by first unpacking the notions of copyright built into NAICS and then analyzing how open source software production is rendered visible, or invisible, within the structure of NAICS. Doing so foregrounds the theoretical and ideological commitments embedded in NAICS, a classification scheme whose foundational principles are considered rational and consistent.

2. The New Economy And NAICS

The last years of the twentieth century were characterized by an almost giddy optimism about the beneficent impact of information technologies in bringing the world to a new and higher plateau of production and consumption. The large number of companies centering their business models on the Internet best exemplified the era. Often called "The New Economy," the era was and continues to be characterized by a sense that information technologies have been ushering in a world where the costs of production are reduced, distribution networks are dramatically enlarged, and consumers have instant access to many new products and services. The New Economy represents the age of "friction-free capitalism," in the phrase of Bill Gates (1996), where consumers and producers are matched without the need for cumbersome middlemen, an age of radical disintermediation.

Not surprisingly the perceived advent of the New Economy suggested a need to measure its vaunted productivity gains and to account for the new products and services that characterize it. In the United States this need was crystallized in the call for and eventual construction of a new industrial classification system to be used as the structure for gathering information about economic activity. The American government is required by law to conduct an economic census every five years and the data gathered during the economic census to be organized into categories, or classes, for aggregation and retrieval (U.S. Census Bureau, 2000). The new classification system introduced for use with the 1997 Economic Census, NAICS, replaced the Standard Industrial Classification (SIC) that was first implemented during the Roosevelt Administration. The SIC system was perceived as flawed in many respects, particularly in its inability to capture the emerging industries associated with the New Economy and in measuring the productivity of an economy increasingly driven by services rather than manufacturing. SIC was also found wanting in the lack of conceptual rigor used to categorize industries as they developed and changed. NAICS was created with the goal of remedying these deficits. NAICS's creators took on a daunting task when they decided to abandon the aging and unwieldy SIC in favor of building a classification scheme from scratch. They chose four principles upon which to base their system so that it would not be subject to the inconsistencies that plagued SIC over time. These guiding principles include:

1. NAICS is erected on a production-oriented conceptual framework. This means that producing units that use the same or similar production processes are grouped together in NAICS.
2. NAICS gives special attention to developing production-oriented classifications for (a) new and emerging industries, (b) service industries in general, and (c) industries engaged in the production of advanced technologies.
3. Time series continuity is maintained to the extent possible.
4. The system strives for compatibility with the two-digit level of the International Standard Industrial Classification of All Economic Activities (ISIC Rev. 3) of the United Nations. (U.S. Office of Management and Budget, 2005)

NAICS added some 350 industries that SIC did not recognize in a structure that begins with two-digit numbers for 20 major broad industries, called sectors, and works down to six-digit numbers denoting small sub-industries within those sectors. Figure 1 excerpts selected sub-industries in Sector 51.

2002 NAICS Code	*2002 NAICS Title*
51	Information
511	Publishing Industries (except Internet)
5111	Newspaper, Periodical, Book, and Directory Publishers
51112	Periodical Publishers
51113	Book Publishers
51114	Directory and Mailing List Publishers
5112	Software Publishers
515	Broadcasting (except Internet)
5151	Radio and Television Broadcasting
51511	Radio Broadcasting
51512	Television Broadcasting
5152	Cable and Other Subscription Programming
516	Internet Publishing and Broadcasting
518	Internet Service Providers, Web Search Portals, and Data Processing Services
5181	Internet Service Providers and Web Search Portals
51811	Internet Service Providers and Web Search Portals
518111	Internet Service Providers
518112	Web Search Portals
5182	Data Processing, Hosting, and Related Services

Fig 1. *NAICS Information Sector 51 with selected sub-industries.*
Source: U.S. Census Bureau, 2002 *NAICS Codes and Titles*, http://www.census.gov/epcd/naics02/naicod02.htm#N51

The U.S. government's news release announcing the creation of the classification scheme included a declaration from an Office of Management and Budget (OMB) official that "NAICS is a bold response to those who say that Federal statistics on business overlook new and important economic activities. This new system will ensure that business and public decisions can reflect real economic changes better and sooner" (U.S. Office of Management and Budget, 1997). A year later, the Economic Classification Policy Committee (1998), whose members had been appointed by OMB, issued a colorful brochure about the system titled *NAICS: New Data for a New Economy*.

The new system took several years to develop. A central way in which the creators of the new classification sought to enhance conceptual clarity was to commit to the aggregation of economic activity under the principle of like processes of production. This was one of the four principles that guided the construction of NAICS, and it was key. As Jack Triplett, the chief architect of NAICS, has explained it, "NAICS was the first industry (or other) classification system to be erected on an economic concept. In NAICS, establishments (productive units) are grouped into industries on the basis of similarities in their production processes." (Triplett, 2002) This narrative about the need for change, from the irrational SIC system of the past to the rational NAICS of the New Economy, follows well-worn narrative stages where new circumstances, in this case forms of economic activity and products, call for new ways of classifying and measuring that activity. NAICS fits the narrative scheme well, and we have previously highlighted the importance of orthodox economic theory in that history as it relates to NAICS

(Elichirigoity & Malone, 2002). What may not be clear from the traditional historical narrative is that economic classification systems co-emerge with the objects and practices of activity that they purport to measure. As we argued previously (Malone & Elichirigoity, 2003) "information" as a product did not exist within SIC, and thus it could not be measured as such. NAICS provides a classificatory space where "information" becomes a measurable product. In this sense, then, a classification system, like a train on a track, cannot go beyond what is laid out as its conceptual framework. NAICS, as well as any other form of statistical classification, needs to be understood as emerging conjointly with the field it claims to measure, in this case the North American economies. Classification systems make visible the fields of intervention and measurement they refer to. This characteristic of taxonomic work is, of course, not new. As Bowker and Star (1999) have persuasively argued, classification of work is inherently social and political. The economic activities NAICS measures are the activities the very measurement of which makes visible.

The classification system is the filter through which the area of activity known as the "economy" is seen. Apply a different filter, such as SIC, and different conceptions of the economy can appear or disappear. For example, NAICS does a rather poor job of organizing data about the nonprofit sector, an enormous area of activity, especially in the U.S., but also in Canada. As Lampkin, Romeo and Finnin (2001) argue, "Because NAICS was created on a strict production paradigm, it is difficult to classify organizations that do not have a strict economic output" (p. 787). NAICS also fails to categorize and thus make measurable areas of economic activity such as underground barter, babysitting, and "housework that supports the economic development of the household" (Boettcher, 1999, p. 6). Such shortcomings may represent what Beghtol (2003) has referred to as a "naïve classification system," one designed for a particular purpose by a subject matter expert with no interest in the discipline of classification. They may also be construed as a symptom of the specific theoretical commitments that are embedded in NAICS. These theoretical commitments precede the functioning of NAICS, as Campbell (2003) argues, by "enforcing standardized data collection" (p.232) across national borders. Thus the apparently context-free statistics that are made possible by NAICS hide but do not eliminate the theoretical commitments that we foreground in this paper and in our previous research (Malone & Elichirigoity, 2003; Elichirigoity & Malone, 2002).

3. NAICS, Open Source And Copyright

"Many of the industries in the NAICS Information sector are engaged in producing products protected by copyright law, or in distributing them," states the NAICS definition of Sector 51 Information (U.S. Census Bureau, n.d.). "Examples are traditional publishing industries, software and directory and mailing list publishing industries, and film and sound industries." While open source is not specifically mentioned, the NAICS definition of Sector 51 touches on two factors related to both proprietary and open source software production. Firstly, it involves the uniqueness of information as an often intangible third thing between good and service. Secondly, it involves the ease with which information can be recombined or repackaged to create additional revenue. A closer look at each of these in turn reveals the underlying understanding of copyright as part of the software production process.

The intangibility of information products makes it easy to violate the copyright holder's exclusive right to copy and distribute the work. "Most of these [information sector] products are protected from unlawful reproduction by copyright laws," the definition states. "The intangible property aspect of information and cultural products makes the processes involved in their production and distribution very different from goods and services. Only those possessing the rights to these works are authorized to reproduce, alter, improve, and distribute them. Acquiring and using these rights often involves significant costs" (U.S. Census Bureau, n.d.). The language in this part of the sector definition suggests a particular view of copyright, one drawn more from the perspective of the Motion Picture Association of America than from the independent documentary filmmaker and more from the Recording Industry Association of America than from your local garage band. In U.S. copyright law, the creator of original work owns the copyright to it as soon as it is in fixed form. In the NAICS definition, rights are things to be acquired from others, at some cost to the acquirer but with a return on investment with the right sorts of uses. The costs of acquiring exclusive rights and of defending those rights against movie, music, and software piracy can indeed be expensive, but the potential for profit is great. The activity involved in securing and defending copyright is part of the pro-

duction process represented in the first of the four guiding principles listed above.

Those who go to the trouble of acquiring copyright through copyright transfer agreements signed with the original creators of the work or who acquire only some rights through licensing agreements with copyright holders also invest in ways to use the material, often through recombining or repackaging the information. The combination of different information products, the addition of a good or service to an information product, the rearrangement and branding of a information product all represented value added, and the process of adding that value represents a productive activity that can generate measurable data. The NAICS definition cites as one example a distributor who charges for advertisers to have their information included in a product (such as a televised sitcom) that is distributed at no charge to end-users and as another example the database producer who pays for the rights to newspaper and magazine articles, adds an interface and search engine, and charges end-users for access. Even within these investments in use, there is an investment in rights management. Without the prior and continuing investment in copyright acquisition and maintenance, the investment in use (that is, in producing products that use the rights acquired and maintained) is pointless.

Clearly, the NAICS definition rests on a traditional business model in which incentives must be present for innovation to occur. In this model, the exclusive rights of copyright holders function as the incentives. (Weber, 2004, pp. 190-194) Defenders of the traditional business model have argued its necessity on technical, commercial, and legal grounds, according to Open Source Initiative General Counsel Lawrence Rosen (2005). Rosen summarizes the traditionalists' arguments this way: "No free or open source project ... could develop the highly complex and robust code necessary for modern software applications ... No free or open source project could survive commercially, given the high costs of quality programming, and the inability to exclude others from the benefits of that quality ..." (pp. xv-xvi).

Information Sector companies that have invested in acquiring and maintaining exclusive rights to intellectual property and in creating products and services based on those exclusive rights in the hopes of earning profits are construed as productive units. What we may lose sight of, however, is that the processes they engage in may actually run counter to productivity and that there may be other productive units unaccounted for within such a definition.

3.1 Copyright as Anti-productive

An aspect that is invisible in NAICS is the cost (which we could define as negative productivity), to individual producers and to the economy as a whole, of organizing economic production around a narrow definition of copyright. Thus, for example, current militant notions of copyright protection have vastly increased the cost of producing documentaries. As Aufderheide (2005) asserts, "unbalanced interpretation of copyright leads to a creative stranglehold." She illustrates her point with three examples where the cost of producing a documentary increased dramatically because of new interpretations of copyright. In one case filmmaker David Van Taylor had to pay the estate of Irving Berlin several thousand dollars when the object of his documentary, Oliver North, started interacting with a large audience that was singing Berlin's "God Bless America." While we certainly understand the complexity of the arguments about the value and sanctity of copyright we note that NAICS would register this added cost as an increase in cultural productivity, while the real effect is one of chilling that production. By hiking the measuring rod to an ideologically specific notion of copyright, NAICS serves to obscure certain economic measurements that another classification system or another approach could potentially make visible.

The intensification of a narrow conception of intellectual property is starting to have deleterious effects in other areas such as biotechnology, a quintessential New Economy industry, through what is called "the tragedy of the anticommons" (Heller, 1998). The term was first coined by Michael Heller to describe the multiplicity of ownership claims to property in post-Soviet Russia, often resulting in not one owner being able to develop the property. This phenomenon comes into existence when there are too many individuals or corporations having rights of exclusion to a rare or scarce resource. This is increasingly common in New Economy industries where too many different owners own specific techniques or paths of software development, in fragmented elements. The tragedy occurs when the many owners, by acting according to rational economic theory, exclude others to maximize their control and profits. The end result, by each owner acting in this way, is a collective devaluation and under-utilization of a specific set of resources. The notion has been extended to biomedical research. For example, Heller and Eisenberg (1998) show that patents to newly identified DNA sequence fragments have increased the complexity of

research for therapeutic needs. As they argue, "commercial products, such as therapeutic proteins or genetic diagnostic tests, are more likely to require the use of multiple fragments. A proliferation of patents on individual fragments held by different owners seems inevitably to require costly future transactions to bundle licenses together before a firm can have an effective right to develop these products" (Heller & Eisenberg, 1998, p. 699). Benkler (2001) argues regarding the information industry that "expansion of exclusive private rights in information tilts the institutional ecosystem within which information is produced against peer production and in favor of industrial production..." (p. 86). Boyle (2003) concludes, "More property rights, even though they supposedly offer greater incentives, do not necessarily make for more and better production and innovation – sometimes just the opposite is true. It may be that intellectual property rights *slow down* innovation, by putting multiple roadblocks, multiple necessary licenses, in the way of subsequent innovations" (p. 44 [emphasis in the original]). In the NAICS discussion of copyright is an implicit definition that emphasizes protection rather than distribution or use, even when such protection may impede productivity.

3.2 Distributed Productive Units

NAICS rests on the comforting but increasingly inaccurate notion of the permanent character and identity of economic actors. It assumes that the business establishments whose data it organizes and measures are largely unchanging. Thus the notion of "productive unit" in NAICS looks just like the establishments that were characterized and measured under the old regime of SIC. This tacit naturalization of what constitutes an economic actor hides the emergence of new actors who do not neatly fall within the categories assumed by NAICS.

For example, there's no reason why a virtual community of open source software developers cannot be construed as a "productive unit." Programmers who download open source code, fix buggy and inelegant code, and submit the revised code to the project leader for acceptance and further distribution, use essentially the same practices and procedures as a conventional business establishment's software development team. Practically, however, the virtual community's productive labors cannot be captured and represented in the U.S. Economic Census, the main deployment that organizes data by NAICS codes and categories. Business establishments are surveyed for the Economic Census, which reports such measures of value and productivity as number of employees and amount of revenue. To supplement the survey data, Economic Census staff fills in with information from pre-existing federal government files and records. Until the open source community has a permanent address in the U.S., hires employees for pay, collects revenue, and files tax returns, as conventional business establishments do, economic data about it will remain unreported, at least by the Economic Census. The Economic Census is not NAICS, of course. But it does help highlight a theory-into-practice understanding of what happens as NAICS is applied to the problem of organizing economic data when the organization of economic production is dynamic.

3.3 Open Source as a New Mode of Production

At about the same time that NAICS was being developed and revised in the 1990s, a novel economic mode of organizing production was emerging with the free software phenomenon evolving into the open source movement and demonstrating that innovation and economic productivity could take place within a model of collaboration rather than competition. Glyn Moody's (2001) popular account of Linus Torvalds's efforts to coordinate a volunteer army of programmers and users to create the Linux operating system describes the work processes as they evolved into a new mode of production." ... Linux was not just a piece of software but an entire development methodology ... using a distributed team, connected by the Internet, to feed through smaller elements that together make up the whole ... and to let users debug the system" (Moody, 2001, p. 99).

One central innovation characteristic of the free and open source phenomenon was the creation of licensing agreements such as Richard Stallman's widely adopted General Public License (GPL), which gives away the copyright holder's exclusive rights to the source code on condition that any recipient or user who modifies the code must issue it under the same GPL. In other words, signatories of the GPL have the freedom to use the software, have access to the code and change it if desired, and distribute the software and any changes made to it. Source code thus is not bounded by the protective enclosure of exclusive rights. It is open, and open to being copied, distributed, and revised, infinitely. Weber (2004) sums up the significance of the GPL: "Stallman cleverly inverted copyright to something

he called 'copyleft,' which in effect reverses the logic of keeping software source code secret. Instead, the GPL uses copyright law to ensure that free software *and derivative works from free software* remain free" (p. 48 [emphasis in the original]).

The novel use of the licensing concept in open source software production is coupled with novel forms of production and work coordination. Traditionally, and in mainstream economic theory, production is organized around market pricing and/or corporate structures. The question of why not all transactions are market transactions was a question discussed by institutional economist Ronald Coase. In his seminal 1937 article titled "The Nature of the Firm," Coase argues that corporate structures organized on the basis of managerial command and control systems emerge as the result of increasing transaction costs resulting from economic exchange, such as having to pay intermediaries or having to enforce terms of contract. (Coase, 1937) At some point the costs are higher than if the individual entrepreneur engaged in an economic activity hires other people to lower transactions costs. Yochai Benkler argues that open source production, which he calls "commons-based peer production," represents a third mode of production that is neither a pure market process nor a corporate-based process (Benkler, 2005, p.169). He further states: "This emerging third model is distinct from the other two, and has certain systematic advantages over the other two in identifying and allocating human capital/creativity" (p.171).

He goes on to say that this new mode of production and coordination of human capital is enhanced by widespread availability of information and communication networks. Commons-based peer production, according to Benkler, benefits from the ability of the system to let individuals choose what creative possibilities they want to engage in and at what level of commitment, thus facilitating the maximization of creative resources available from large pools of qualified individuals (p.178). Furthermore, the system also benefits from decentralized but efficient coordination of the final product resulting from several processes that include norm-based social organization and embedded solutions in the collaboration platform itself (p. 190). In other words, open source is a mode of production uniquely fit to economic activities that are largely based on human capital, activities often associated with the New Economy.

Because NAICS is committed, through its reliance on copyright, to notions of economic activity that only fit the two traditional forms of economic organization explained by Coase, the taxonomy cannot make room to measure activity occurring as commons-based peer production.

Despite their coincident development, NAICS does not take into account open source production. As we previously discussed it (Malone & Elichirigoity, 2003), one of the new NAICS categories, Sector 51 Information, did a good job of making visible and measurable specific conceptualizations of the communication and information services that had grown and multiplied remarkably since the final revision of SIC in 1987. However, the open source software industry remains largely invisible and unmeasurable within Sector 51 and its subcategories even now. This phenomenon is, paradoxically, exacerbated by what Bowker and Star (2001) call the "imaginary of unlimited countability," (p. 424) the notion that everything can be counted. In the case of NAICS this leads to an over-commitment to measure commodities, or "information as thing," (Buckland, 1991) rather than measuring processes, which are often difficult to quantify, but increasingly important, forms of economic activity. NAICS, then, fails to represent the New Economy emerging as open source production at two levels: a) it cannot account for the immense productive capacity of the social ecology of open source and b) it cannot see the process itself, thus obscuring one of the more unique features of the New Economy.

Lest there be any doubt that open source is a significant economic development worthy of documenting, a few statistics are in order. According to its homepage, SourceForge.net, which manages "the largest repository of Open Source code and applications available on the Internet," has almost 100,000 registered projects and more than 1 million registered users. Email and Web users benefit from open source software, large corporations and small companies rely on it for managing their business processes, and computer industry giants have embraced open source software development, including none other than IBM, which claims to have invested more than $1 billion in Linux development alone (IBM, n.d.).

One of the shortcomings that NAICS's predecessor SIC was criticized for – its inadequacy in accounting for the new communications and information industries – seems to affect NAICS as well. The inadequacy springs in part from two specific characteristics of open source production. First, there are some things about open source production that make it difficult to identify as an economic activity

about which data can be collected. Production of open source code involves tacit rules and practices that may not be clear to outsiders and that may only recently, through published essays, articles, and books, have become clear to insiders who produce, share, test, and revise code. Further, open source code is essentially intangible since it is uploaded and downloaded via the Internet, often across national borders, and worked on while in computer memory. Intangibles are inherently difficult to observe, capture, and classify. As we focus on NAICS itself, we find that even this disciplined rational classification scheme relies on assumptions and assertions that undermine its own stated principles. The NAICS assumptions and assertions about copyright as a production input for industries classed in Sector 51 Information are at odds with the open source movement's copyleft approach to intellectual property. Weber (2004, 1) captures the essential difference between proprietary and open source software:

> The conventional notion of property is, of course, the right to exclude you from using something that belongs to me. Property in open source is configured fundamentally around the right to distribute, not the right to exclude. If that sentence feels awkward on first reading, that is a testimony to just how deeply embedded in our intuitions and institutions the exclusion view of property really is.

In the NAICS definition of Sector 51, distribution is where the risk of losing exclusive rights is the greatest because with distribution of intangible information products comes a diminution of control over copying, deriving new products, and redistributing without authorization. In contrast, distribution is the lifeblood of open source production. Yet, in NAICS the production of open source software is no different from the production of proprietary software, rendering open source invisible as a separately identifiable and measurable industry.

3.4. Classifying Red Hat

What follows is a specific example of how NAICS obscures some aspects of economic activity related to open source and the necessary distinctions and novel forms of production that could perhaps be measured within a taxonomic system embodying different notions of production and consumption. For our example we look at a company named Red Hat. This company creates value-added by customizing the Linux operating system to serve businesses' needs and by bundling it with services that non-programmers, who are not interested in looking at or changing Linux's code, are happy to buy. Red Hat is a pioneer that proved that money could be made from the new business model that open source code creation and distribution represent (Moody, 2001, pp. 96-98). To the extent that Red Hat follows a course recognized by NAICS – adding value by combining and/or repackaging information goods and services – its production processes can be comprehended and assigned to a subcategory of industries with like production processes. And, indeed, they have been.

The Red Hat, Inc. (2005) company profile in *Hoover's Company Records* lists these:

NAICS CODES:
511210 – Software Publishers
518210 – Data Processing, Hosting, and Related Services
54151 – Computer Systems Design and Related Services
541511 – Custom Computer Programming Services
541512 – Computer Systems Design Services
541519 – Other Computer Related Services
611420 – Computer Training

Looked at in this way, Red Hat's production process is not so different from those of any other company that acquires rights to proprietary software and bundles it with services; the assumption is that they should not be in separate categories. Because the category for open source production is not there, we do not miss it.

But let's look more closely at what this means in practice. The first two NAICS codes assigned to Red Hat in *Hoover's Company Records* are 511210 – Software Publishers and 518210 – Data Processing, Hosting, and Related Services. If one searches in Hoover's for other companies that also operate in those NAICS classes, one will retrieve more than 350 results. Among them is Microsoft, which zealously guards its own proprietary code and disparages open source. Since the open-source Linux operating system that Red Hat has helped build an entire industry around was originally developed to offer an alternative to Microsoft's closed operating system, it is ironic that NAICS sees Microsoft and Red Hat as more alike than different, at least when it comes to

the software publishing and processing and hosting services aspects of their business.

Taken from another perspective, consider the data reported in the latest U.S. Economic Census (2002) for the software publishing sub-sector of the Information Sector, NAICS code 5112. (The 2002 Economic Census does not break these codes down beyond the four-digit level.) In 2002, there were 9,899 establishments doing business in this sub-sector with revenues of $103,737,152 and paid employees numbering 353,285. Somewhere in those figures presumably are both Red Hat and Microsoft. If one's interest lies in understanding the impact of open source products and services in the new economy, having the data categorized the way NAICS does it will not help.

If the researcher is interested instead in two other Information sub-sectors, the publishing and broadcasting industries, NAICS takes a different approach, separating "Internet publishing and broadcasting" (NAICS code 516) from "Publishing (except Internet)" (511) and "Broadcasting (except Internet)" (515). Separating out the Internet-based sub-sector in 2002 was a change from the 1997 NAICS scheme, made, at least in part, to satisfy the second of the four principles, to demarcate emerging industries and those involved in advanced technologies.

In the case of integrating open source software production with proprietary software production, it is difficult to fathom which of the four principles is at work. From the Red Hat example, it's clear that the kind of business Red Hat is in existed in the days of the SIC, so it's not a case of an emerging new industry subcategory. Still working with the Red Hat example, it's not clear that the company's production processes are like those of a company that bundles services with propriety software and markets the combined service and good. On the surface, they seem quite similar. But in the Information Sector as NAICS defines it, the production processes involving the investment in copyright acquisition and maintenance loom large. In an open source company like Red Hat, copyright acquisition and maintenance processes are simplified into a copyleft reciprocal agreement. A proprietary software company that invests in protecting its source code from discovery, developing technical barriers to piracy, retaining staff attorneys to guard its rights, and mounting a public campaign of education and/or persuasion to dissuade its customers and the public from making illegal copies would appear to have a different production process from a company that signs a simple GPL-type license and then proceeds to concentrate on developing products and services and marketing and distributing them.

4. Conclusion

Our critical reading of NAICS and the claims of its authors to be a knowledge organization tool capable of accounting for the New Economy lead to a number of observations. One of them is that the theory and the actual production of taxonomies have to include a process of self-reflexivity regarding the theoretical commitments embedded in the process of producing the taxonomy. In the case of NAICS, some economic assumptions were taken for granted and left unexamined. Neo-liberal economic theory was assumed to be universal and ahistorical, robbing the taxonomy of the necessary plasticity to accommodate new forms of production to be described and measured.

Another observation has to do with the commitment to one universal principle. As Olson (1996) has clearly shown, adherence to universality can result in "marginalization of the unconventional." In the case of NAICS, the universal principle was one of similarity in production processes as a way to aggregate information about economic activity. Although the goal of clarity is admirable, it is not well-served by organizing complex taxonomies. The commitment to this one principle coupled with unexamined assumptions behind it can lead to a rigidity of categories. In an era where technological processes such as digitization and genetic engineering are giving rise to large number of hybrids that escape the boundaries of modernist taxonomies, the possibility of universal purity should be given up in favor of flexible categorization regimes. In this respect open source production presents a particularly interesting range of taxonomic possibilities not capturable within one underlying principle. As anthropologist Christopher Kelty (2004, p. 501) suggests, for many activists, often in countries other than the United States, open source software development is not just a novel form of organizing production but also a form of politics and critique of mainstream economic practices. To some extent, open source becomes a form of protest or freedom of speech involving a range of practices and meanings not immediately divisible between economic and political activity and not easily classifiable under a universalist umbrella.

The adoption of a narrow understanding of copyright in the definition of the NAICS Information

Sector points to a taken-for-granted aspect of the classification system, namely, its embeddedness in a much larger legal and political infrastructure that privileges specific values. The classification system helps enact a specific conception of economic production and value. This infrastructure tends to work against the emergence of new economic arrangements and, when they do emerge, their useful depiction in an economic classification system designed for data organization and aggregation. NAICS, by assuming a historically contingent institutional ecosystem to be natural and objective, is poorly equipped to "see," to allow the conceptualization of measurable economic producers that fall outside the categories sustained by the current legal-political framework. NAICS, then, cannot be understood as the taxonomy of an objective reality of producers and consumers but as an active element of a historically specific framework that makes visible a particular type of producer, consumer, and economy.

References

Aufderheide, P. 2005. The creative copyright stranglehold. *Flow*. Available at: http://jot.communication.utexas.edu/flow/?jot=view&id=701

Beghtol, C. 2003. Classification for information retrieval and classification for knowledge discovery: Relationships between "professional" and "naïve" classifications. *Knowledge organization* 30: 64-73.

Benkler, Y. 2001. The battle over the institutional ecosystem in the digital environment. *Communications of the ACM* 44 (2): 84-90.

Benkler, Y. 2005. Coase's penguin, or, Linux and the nature of the firm. In *Code: Collaborative ownership and the digital economy*, ed. R. A. Ghosh, 169-206. Cambridge, MA: MIT Press.

Boettcher, J. 1999. Challenges and opportunities presented by NAICS. *Journal of business and finance librarianship* 5 (2): 3-13.

Bowker, G.C. and S. L. Star. 1999. *Sorting things out: Classification and its consequences.* Cambridge, MA: MIT Press.

Bowker, G.C. and S. L. Star. 2001. Pure, real and rational numbers: The American imaginary of countability. *Social studies of science* 31: 422-25.

Boyle, J. 2003. The second enclosure movement and the construction of the public domain. *Law and contemporary problems* 66 (Winter/Spring): 33-74. Available at: http://ssrn.com/abstract=470983

Buckland, M. 1991. Information as thing. *Journal of the American Society of Information Science* 42: 351-60.

Campbell, G. 2003. Global abstractions: The classification of international economic data for bibliographic and statistical purposes. *Cataloging & classification quarterly* 37 (1/2): 221-34.

Coase, R. 1937. The nature of the firm. *Economica* 4 (16): 386-405.

Elichirigoity, F. and C. K. Malone. 2002. Representing the global economy: The North American Industry Classification System. In *Challenges in knowledge representation and organization for the 21st century. Integration of Knowledge Across Boundaries: Proceedings of the Seventh International ISKO Conference*, July 10-13, 2002, Granada, Spain, ed. M. J. Lopez-Huertas, 345-350. Würzburg: Ergon Verlag.

Gates, B., N. Myhrvold, and P. Rinearson. 1996. *The road ahead.* New York: Penguin Books.

Heller, M.A. 1998. The tragedy of the anticommons: Property in the transition from Marx to markets. *Harvard law review* 111: 622-88.

Heller, M. A. and R. Eisenberg. 1998. Can patents deter innovation? The anticommons in biomedical research. *Science* 280 (5364): 698-701.

Hoover's. 2005, March 14. Red Hat, Inc. Hoover's Company Records-In-Depth Records. Retrieved April 29, 2005, from Lexis-Nexis Academic Universe.

IBM. New to open source. Available at: http://www-128.ibm.com/developerworks/opensource/newto/.

Kelty, C. 2004. Culture's open sources: Software, copyright, and cultural critique. *Anthropological quarterly* 77: 499-505.

Lampkin, L., S. Romeo and E. Finnin. 2001. Introducing the non-profit program classification system: The taxonomy we've been waiting for. *Nonprofit and voluntary sector quarterly* 30: 781-92.

Malone, C.K. and F. Elichirigoity. 2003. Information as commodity and economic sector: Its emergence in the discourse of industrial classification. *Journal of the American Society for Information Science and Technology* 54: 512-20.

Moody, G. 2001. *Rebel code: The inside story of Linux and the open source revolution.* Cambridge, MA: Perseus Publishing.

OSTG (Open Source Technology Group). SourceForge.net. http://sourceforge.net.

Olson, H. 1996. *Between control and chaos: An ethical perspective on authority control.* In Authority

Control in the 21st Century: An Invitational Conference, March 31 – April 1, 1996: Proceedings. Available at: http://digitalarchive.oclc.org/da/ViewObject.jsp?fileid=0000003520:000000091795&reqid=354.

Red Hat, Inc. (company profile). (2005). In *Hoover's Company Records*. Retrieved April 29, 2005, from Lexis-Nexis Academic Universe.

Rosen, L. 2004. *Open source licensing: Software freedom and intellectual property law.* Upper Saddle River, NJ: Prentice Hall PTR.

SourceForge.net. http://sourceforge.net.

Triplett, J.E. 2002. Industries, products and aggregations: NAICS provision of information for the new economy. Paper prepared for IAOS Meetings, London, August 2002. Available at: http://www.statistics.gov.uk/iaoslondon2002/contributed_papers/downloads/IP_Triplett.doc

U.S. Census Bureau. No Date. *NAICS Sector: 51 Information.* Available at: http://www.census.gov/epcd/ec97/def/51.HTM.

U.S. Census Bureau. 2000. *History of the 1997 economic census.* Washington, DC: U.S. Department of Commerce, Economics and Statistics Administration. Available at: http://www.census.gov/prod/ec97/pol00-hec.pdf

U.S. Economic Census. 2002. Available at: http://www.census.gov/econ/census02/.

U.S. Economic Classification Policy Committee. 1998. *NAICS: New data for a new economy.* Available at: http://www.census.gov/pub/epcd/www/pdf/naicsdat.pdf.

U.S. Office of Management and Budget. 1997. *Administration Introduces new industry classification system.* Available at: http://www.census.gov/epcd/naics/pressrel.html.

U.S. Office of Management and Budget. 2005. North American Industry Classification System – Updates for 2007. *Federal Register,* Available at: http://www.census.gov/epcd/naics07/index.html.

Weber, S. 2004. *The success of open source.* Cambridge, MA: Harvard University Press.