

**PERSPECTIVE**

# Archives of the People, by the People, for the People

Max J. Evans

**Abstract**

Archivists today are caught between an expanding volume of records and a growing public expectation that every page in every document is online and indexed. With so many records and so few resources to provide on-demand access to them, the problem seems intractable. More money alone is not the answer; larger appropriations or donations cannot solve this problem. Instead, archivists must fundamentally shift the way they think about their roles and develop alternative means and methods for doing archival work.

This paper introduces the concept of commons-based peer-production as a means of turning collections inside out. It encourages archival institutions to reinvent themselves, and, in collaboration with other archives and with other types of organizations, to organize archival work in concert with a curious and interested public.

**The Problem**

American archivists today face fundamental challenges to their basic beliefs. They are busy managing an enormous and growing volume of records.<sup>1</sup>

This article is based on papers given at professional meetings of the New England Archivists in Storrs, Conn.; the Society of Ohio Archivists in Columbus, Ohio; the joint meeting of Northwest Archivists/Society of Rocky Mountain Archivists/Conference of Inter-Mountain Archivists in Las Vegas, Nev.; the joint meeting of the New York Archives Conference/Capital Area Archivists of New York/Archivists Round Table of Metropolitan New York in Albany, N.Y.; and at the Henry Ford Museum in Dearborn, Mich. The author thanks all who contributed their ideas and offered encouragement. The author also acknowledges the critical comments of his colleagues Kathleen Williams and Keith Donohue during the preparation of this article. I also express my gratitude to the late Prof. Roy Rosenzweig who listened to an early synopsis of this paper and introduced me to the writings of Yochai Benkler.

<sup>1</sup> U.S. archives and other repositories hold collectively at least 11 million cubic feet of records and historical manuscripts—roughly 33 billion pages. This is a very conservative estimate based on these data: The National Archives holdings of “traditional records” equal 3.3 million cubic feet. National Archives and Records Administration, “Performance and Accountability Report, FY 2006,” 75, available at <http://www.archives.gov/about/plans-reports/performance-accountability/2006/nara-2006-par-complete.pdf>, accessed 11 June 2007. “The combined holdings for all state archives in 2006 amounted

Their methods now meet headlong the rising expectations of an information-hungry citizenry, transformed by the Internet in what must be one of history's most astoundingly rapid adaptations to technology.

The problem, simply stated, is this: Archives<sup>2</sup> barely have the resources to manage existing collections, new accessions, and new and changing forms of records, including those "born digital." Budget cuts and a veritable tsunami of records mean larger backlogs and knowing less, not more, about the collections held. For the archivist, the Information Age means many more records to inventory, appraise, accession, and process. But it suggests to the rest of the world that all information will be easily and quickly available. The Internet promises to increase the public's awareness and use of archives and historical records—a future I think we all want to encourage. But reality intrudes. If this brave new world of Web access to archives depends on document-level description, archivists are doomed by the sheer mass waiting in the unprocessed stacks. The expense of extracting and entering item-by-item metadata multiplies many times over the huge costs involved in scanning entire collections.

It's a conundrum without an obvious solution. But solved it must be, lest the archival profession be consigned to the antiquarian role of Keepers of the Ancient Writings, a guild good at saving information, but not effective at retrieving and revealing it.

#### The Re-engineered Archives

This paper proposes a model for archival activities that systematically reconciles these extremes. Some may see this model as a paradigm shift, but it is no more than a re-articulation of archival principles, refreshed and presented to meet the realities of the information economy and the technological and social climate of the twenty-first century.<sup>3</sup> It is built upon the entirely rational assumption that

to more than 2.7 million cubic feet . . .," *The State of State Records: A Status Report on State Archives and Records Management Programs in the United States* (Council of State Archivists, January 2007), available at <http://www.statearchivists.org/reports/2007-ARMreport/StateARMS-2006rpt-final.pdf>, accessed 8 June 2007. Vicki Walsh extrapolates data from twenty-two states to conclude that there were about 5 million linear feet of nongovernment records nationwide in 1998. *Where History Begins: A Report on Historical Records Repositories in the United States* (Council of State Historical Records Coordinators, 1998), available at <http://www.statearchivists.org/reports/HRRS/HRRSALL.PDF>, accessed 8 June 2007. Email to author from Vicki Walsh, 8 June 2007. There are no good studies of the volume of records held by county and local governments. None of these numbers include the vast and growing volume of nonpaper records: microforms, audio recordings, video tape and motion picture films, and electronic records.

<sup>2</sup> I use the word *archives* here to include archival institutions, historical records repositories, special collections, libraries, and museums that hold archives, personal papers, and manuscripts. See Richard Pearce-Moses, *A Glossary of Archival and Records Terminology* (Chicago: Society of American Archivists, 2005), 30–32, available at <http://www.archivists.org/glossary/index.asp>, accessed 7 June 2007.

<sup>3</sup> "Most innovations are not radical breakthroughs, but creative recombinations of existing ideas." Christine W. Letts, et al. in *High Performance Nonprofit Organizations: Managing Upstream for Greater Impact* (New York: John Wiley and Sons, 1999), 79.

archivists must set priorities and build alliances to be effective in today's information economy. Archivists cannot collect everything and they cannot treat all collections at the same level. Nor can they operate in isolation.

#### *Initial Processing*

After appraisal and collecting, the next priority of any archival institution must be to gain initial legal, intellectual, and physical control through accessioning and basic processing. For each collection, every archives should be obliged to follow the Greene and Meissner<sup>4</sup> model of initial processing, producing and publishing online a suite of basic descriptive tools: a MARC catalog record that summarizes the context, content, and physical location of each collection;<sup>5</sup> and a "container list" as a minimal finding aid marked up in EAD that reveals the arrangement and hierarchy. This description provides meaningful but high-level access points. Such finding aids, found in online repository catalogs and in state, regional, and national or international databases, effectively publicize collections' existence, provided the metadata are exposed to the Web. This standard, if accepted and followed, would effectively expose hidden collections.<sup>6</sup> But this model requires a shift in values. Instead of thoroughly processing only a few collections while the vast majority occupies that purgatory called the "backlog," archivists must apply the principle that each and every collection be made known. The archives of the future depend upon redesigned systems that incorporate lean<sup>7</sup> and effective methods of processing and describing archival holdings, eliminate backlogs, and create high-level, but still useful, discovery tools for customers. "More product"—more collections processed to the minimum appropriate level—will improve access, especially if archives publish online collection-level descriptions and, when appropriate, companion archival finding aids. Basic description sets the stage for establishing additional priorities.

<sup>4</sup> Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist* 68 (Fall/Winter 2005): 208–63.

<sup>5</sup> By *collection* I mean to include, for public records and other institutional archives, *record series*. I use the more inclusive term for brevity.

<sup>6</sup> This phrase comes from the 2003 Exposing Hidden Collections Conference sponsored by the Association for Research Libraries and held at the Library of Congress. See [http://www.arl.org/rtl/specoll/hidden/EHC\\_conference\\_summary.shtml](http://www.arl.org/rtl/specoll/hidden/EHC_conference_summary.shtml), accessed 11 June 2007.

<sup>7</sup> The word *lean* is used intentionally to refer to a discipline for industrial production of that name. I suggest that archivists can learn from private industries' commitment to reducing waste and unnecessary process steps. The seminal book on this topic is James P. Womack, Daniel T. Jones, and Daniel Roos, *The Machine That Changed the World: How Japan's Secret Weapon in the Global Auto Wars Will Revolutionize Western Industry* (New York: HarperPerennial, 1991). More recent works include Womack and Jones, *Lean Solutions: How Companies and Customers Can Create Value and Wealth Together* (New York: Free Press, 2005) and Mike L. George, *Lean Six Sigma for Service: How to Use Lean Speed and Six Sigma Quality to Improve Services and Transactions* (New York: McGraw-Hill, 2003). The latter work focuses on nonprofit organizations while introducing statistical methods of process improvement.

**Detailed Processing**

Clearly, some collections require more than the minimum processing described by Greene and Meissner. Which collections are these, and how much more processing will they need? These are priority-setting questions. Although archivists familiar with their holdings must certainly be part of the decision-making process, one of the advantages of making all collections known is that researchers' interests and demands create *market forces* that should influence the decisions about additional processing. Indeed, this model argues for a largely demand-driven process that shifts the organization of archival work away from a central, command-and-control model to a more market-oriented approach. It encourages archivists to invite researchers into the decision-making process. What could be more fair and democratic?

To understand customers' demands, archivists must rigorously track the use of collections, but not just to produce aggregate statistics for the annual report. With data about both the nature and the use of collections, together with researchers' comments and requests, archivists can make informed decisions about setting processing priorities, determining which collections should get the fuller treatment of detailed processing.

Initial processing ends with collections arranged and described superficially, usually at the collection/series level or sometimes at the series/subseries level. Detailed processing, however, ends with collections arranged and described at the series and file-unit level (sometimes, but rarely, drilling down to the item level). The descriptive result should be an archival finding aid (marked up in EAD, for reasons that will become clear), published online and linked from the collection-level catalog record. Each finding aid within the repository should also be disclosed, that is, made searchable across all other collections. They should also be made available in state, regional, and national databases (such as the RLG ArchiveGrid). They might also become part of cross-repository, subject-oriented websites.

These detailed finding aids, published online, promote the use of collections. But in a Web 2.0 world, researchers who discover collections and collection components should have several interactive choices: an email address or telephone number by which to contact an archivist to learn more; a way to schedule a visit; or a listing of hours and location so that an unannounced visit can be planned. Or, to move closer to meeting customers' expectations, these detailed finding aids can also become the means to order up archival digitization-on-demand.<sup>8</sup>

<sup>8</sup> The National Archives of Australia has implemented digitization-on-demand. Ted Ling, "Why the Archives Introduced Digitisation on Demand," *RLG DigiNews*, no. 4 (15 August 2002), available at <http://www.rlg.org/preserv/diginews/diginews6-4.html#feature1>, accessed 12 June 2006. Another version of this paper has a more telling title: "Taking It to the Streets: Why the National Archives of Australia Embraced Digitisation on Demand," available at [http://www.naa.gov.au/publications/corporate\\_publications/digitising\\_TLing.pdf](http://www.naa.gov.au/publications/corporate_publications/digitising_TLing.pdf), accessed 12 June 2007.

### **Archival Digitization and Systems**

Archival digitization, or in other words, mass digitization, is the means to digitize *entire archival components* and deliver them online quickly and easily, without the high cost of creating extensive metadata associated with each image. This model contrasts with the digitization of selected documents from a variety of collections to produce digital exhibits or digitizing sample documents to illustrate finding aids. Archival engineering should lead to the invention of Web-based systems in which the EAD-based detailed archival finding aid becomes an online order form. The description of each archival component—series, file, or item<sup>9</sup>—becomes a link to an order form. That form will be prepopulated with descriptive and location information from the EAD instance for that component. The user completes the form with order information.

Archives employees then process the submitted orders. Depending on the nature of the request and an archives' policies, all or part of the component may be digitized. The request may be placed in a queue for later processing, or, if the requester is willing to provide a credit card, the order can be filled more quickly. The first user to pay supports each researcher to follow. All users become part of a community, each contributing to and benefiting from enhanced access to the historical record. Many archives handle photo orders in a similar way: the first user to place the order pays the cost of making the negative. Subsequent requesters are charged only the cost of making prints.

Go back to the finding aid. When researchers next click on the link, instead of an order form, they see digital images of the component. For example, clicking on a finding aid entry for a file unit will open a virtual folder, beginning with the first page of the first item. Navigation buttons and menus allow movement among pages and items. There is no description of each item; like researching among the originals in the reading room, what you see, in the context of the whole, is what you get.

In many ways, however, this surpasses the reading room experience. An online researcher avoids travel and can work outside of reading room hours, conducting research in pajamas at two A.M., if desired. The archival institution benefits, too. Originals will not be mishandled, misfiled, or stolen. Virtual access eliminates box retrieval and may reduce reference consultation.

The processes for achieving this must be carefully engineered and built upon institutional policies. On the front end, reference consultation will be required to determine exactly what to digitize if only part of a component is requested. These parts may then become new components to be incorporated back into the finding aid. The front-end processes must also develop ways to

<sup>9</sup> Any of these components may vary in volume. Some series may be small enough to handle in a single order, while some files or items may be too large. As discussed later, an engineering solution must accommodate these variations by inserting human judgment and negotiation.

estimate the cost of digitizing a component, especially if the requestor is to be charged for all or part.

Digitizing should be done according to standards that assure that the quality is appropriate to the items digitized. A process must be engineered for extracting component data, along with all of its inherited data, from the EAD file. This, along with technical information, then becomes the metadata associated with all images digitized for that component. Other elements required to facilitate navigation and support digital preservation will be added to this metadata.

This limited metadata will not guarantee the precise discovery of each image or document, though words and phrases in the finding aid that describe the component may be used to discover aggregate bodies of records and their online images. This is the archival approach; archives rarely provide item-level description because of the high cost of doing so. This approach likewise does not provide an item description; it does better by providing the document itself. Is the picture of a document worth the thousand words required to describe it? Yes, and more, as digitization with minimal metadata costs much less than digitization with item-level metadata. This is why a request to digitize a part of a component might best be met by digitizing instead the entire component.

Such digitizing-on-demand for researchers who use the archival finding aid as an electronic menu is only one of many ways to prioritize digitization. Another is to digitize in response to internal institutional demands. Archives should engineer systems that digitize any document as it is copied, for whatever purpose—for a researcher in the reading room, for an archival exhibit, as part of making a preservation microfilm, or to fill a Freedom of Information Act request (or a state's equivalent), for example. With digital photocopy machines and hybrid microfilming systems, this should not be difficult, and, in fact, is part of ongoing processes in legal and other business enterprises. These machines must be used with an eye to preserving the originals, but the real challenge will be to engineer systems to track what has been done and to make sure that the images are stored and made accessible through links from the online archival finding aids.

In addition, archival repositories or outside institutions may initiate large-scale digitization projects for their own purposes. Some may be supported by grant funds, others by the commercial interests of the outside organization (such as Google or Ancestry.com), or still others by the motives of nonprofit organizations (such as the Genealogical Society of Utah). In each of these cases, the same digitization processes should be followed: adherence to appropriate standards, repurposing EAD data to produce minimum metadata, and systematically linking bodies of images to the component descriptions in the finding aids.

Engineering these new systems involves tasks that will require the collective wisdom and insights of the archival community. They might be best carried out in an open development environment, in much the same way the Archivists Toolkit is now being developed, or as archivists developed MARC-AMC and

EAD. This work is not all software and standards, however, it also involves industry-style process and work-flow changes within institutions. These changes may be built upon emerging professional best practices, but must accommodate local and institutional variations in staffing, resources, and policies.

#### ***Item-Level Description***

So far, this archival processing model emphasizes minimum description for all collections, more detailed description—but not item-by-item—for selected collections (determined largely by market forces), and digitization of a smaller subset of these collections. Digitization should be seen as a carefully planned response to both external and internal demands. Furthermore, digitization, without extensive metadata, is a means to avoid costly item-level descriptions. With minimum metadata, the collection and its components can be located through the archival discovery system, expanding the standard provenance-based hierarchical method.<sup>10</sup> Provenance-based discovery systems rely upon knowledge of institutional history, functions, and activities. Researchers use this knowledge while *browsing* the archival finding aids to locate relevant sources. Making the content of this hierarchically presented data subject to search enhances the provenance method.

Item-level description has the important role of adding more detailed search terms than an archival finding aid normally contains. Although item-level description is of little value once a researcher finds a specific document, it may be difficult to locate documents *precisely* in collections, even those digitized and published online, without it. Many researchers would like to be able to search for the exact document they need, using the descriptive information associated with the scanned document. The question now, however, is not just one of priorities, but also of responsibilities and obligations. As for priorities; only a very wealthy (and rare) archival institution can afford the cost of describing and indexing every document in its collections without ignoring other collections and failing to process new accessions.

As a practical matter, archivists alone cannot do most systematic, item-level, metadata collection.<sup>11</sup> Indeed, it should not even be their responsibility. Instead,

<sup>10</sup> The provenance-based method of describing archives receives plenty of criticism as being not sufficiently specific and requiring a knowledge of governmental (and other institutional) organization and functions not possessed by most potential researchers. See Richard H. Lyle, "Intellectual Access to Archives," *American Archivist* 43 (Spring/Summer 1980): 64–75, 191–207. Archivists struggle to provide subject and name access as a supplement to the provenance method using added entries in catalog records and by developing indexes. Online searching within and across archival finding aids reduces the need to navigate through the hierarchical structure. Word searching of the text of discovery tools does provide the kind of keyword-in-context anticipated by SPINDEX some forty years ago.

<sup>11</sup> The exception is the quantity of contemporary documentation that is printed or typewritten and may be read and converted to character code by optical character readers (OCR). See, for example, the Governor Leavitt 2K2 Program records (Utah 2002 Winter Olympic Games) at the Utah State Archives, available at <http://historyresearch.utah.gov/digital/26017.htm>, accessed 7 June 2007.

archivists must continue to work with archival aggregations—with the forest, not the trees. They must, however, also organize and facilitate item-level describing and indexing projects. Indeed, unless they do, others will step up and produce their own indexes, and the archives will lose control over improved access to its own holdings. As organizers and facilitators, archivists can recruit, train, and manage a corps of volunteers to index their collections. This is not a new idea. Archives large and small rely on volunteers, such as the many who indexed the Freedman Bureau records at the National Archives or the one, retired legal secretary who indexed a large collection of glass plate negatives at the Utah State Historical Society.<sup>12</sup>

The strategic model for organizing archival work is new, however. It functions within the framework of traditional archival methods, including delivering document images without extensive metadata. But this model adds another dimension based on commons-based peer production.

#### The Commons and the Archives

Underlying the new model for archival work is a view of archives as a common and public good rather than as the protected property of an institution. Levels of intellectual<sup>13</sup> access to the records in archives have been at the discretion of the archivist (or the managers of the institution), based largely on the extent resources can be devoted to the work. Less frequently, public demand drives it. But in the commons-based system, the *users* determine the level of intellectual access.

The *commons* concept developed centuries ago, but has arisen in contemporary social theory discourse as a way of explaining the evolution of new modes of communication—the Internet, in particular—through the cooperation of many interested parties. In a 2002 article published in the *Yale Law Journal*, New York University law professor Yochai Benkler<sup>14</sup> writes of a

<sup>12</sup> The “more than 7,000 historical societies, libraries, museums, academic institutions and other organizations and groups who hold historical records in the United States . . .” benefit from volunteers who “contribute more than 17 million hours of labor to the cause of preserving our documentary history” and in “historical societies, unpaid volunteers outnumber paid professional staff by a ratio of 5 to 1.” Vicki Walsh, *Where History Begins: A Report on Historical Records Repositories in the United States* (Council of State Historical Records Coordinators, 1998), available at <http://www.statearchivists.org/reports/HRRS/HRRSALL.PDF>, accessed 8 June 2007.

<sup>13</sup> I use “access” throughout this paper to mean intellectual access because this is my focus. I acknowledge that there are critical distinctions among legal, intellectual, and physical access. Legal access is a matter of law in the case of federal, state, and local archives. In most other archives, legal access is a matter of ethics, donor agreements, physical condition, and institutional policy. See Mary Jo Pugh, *Providing Reference Services for Archives and Manuscripts* (Chicago: Society of American Archivists, 2005) for a more thorough discussion.

<sup>14</sup> Yochai Benkler, “Coase’s Penguin, or Linux and the Nature of the Firm,” *Yale Law Journal* 112 (New Haven, Yale University Press, 2006), 369–446 and Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven: Yale University Press, 2006). Benkler is now a professor of law at the Yale Law School.

fifteen-year-old social-economic phenomenon in the software development world. This phenomenon, called free software or open source software, involves thousands or even tens of thousands of programmers contributing to large and small scale project[s], where the central organizing principle is that the software remains free of most constraints on copying and use common to proprietary materials. No one “owns” the software in the traditional sense. . . . The result is the emergence of a vibrant, innovative and productive collaboration, whose participants are not organized in firms and do not choose their projects in response to price signals.<sup>15</sup>

Free software, Benkler continues, is “only one example of a much broader social-economic phenomenon, . . . the broad and deep emergence of a new, third mode of production in the digitally networked environment.” He calls this mode “‘commons-based peer-production,’ to distinguish it from the property-and contract-based models of firms and markets. Its central characteristic is that *groups of individuals successfully collaborate on large-scale projects* following a diverse cluster of motivational drives and social signals, rather than either market prices or managerial commands.”<sup>16</sup>

Historically, archives have tended to operate as hierarchical institutions organizing their work around internal demands set by managerial commands. Decisions on everything from acquisitions to access are based in part on the institutional mission and its policies, but guided by the archivists, either through a sense of the relative value of kinds of records or through some conception of users’ interests. Archives can change from this traditional method of managing production by explicitly inviting archival users into the process, using market forces to decide what gets processed in detail. “Archives of the People” are products of customers deciding (voting by their orders) what to digitize.

Those who object to this minimum metadata model should know that creating and publishing digital images with minimum metadata does more than just make records available for research. It also places these images before thousands of potential volunteers who will use new tools for online metadata collection. It is not just minimum metadata; it is *extensible* metadata. The data these volunteers collect may include any combination of comments, controlled- or free-text indexing terms, abstracts, or full-text transcriptions. Each archives’ institutional policies determine the range of choices, as will the methods for recruiting and managing volunteers.

<sup>15</sup> Yochai Benkler, abstract of “Coase’s Penguin,” available at <http://www.benkler.org/CoasesPenguin.html>, accessed 29 August 2005.

<sup>16</sup> Benkler, abstract, emphasis added.

**Archives by the People**

In his article, “Coase’s Penguin, or Linux and the Nature of the Firm,” Benkler describes tens of thousands of individuals who collaborate in five-minute increments to map Mars craters, fulfilling tasks that would normally be performed by PhD astronomers. He writes of a quarter of a million people collaborating to create an important news and commentary site on technology issues; of 25,000 people together creating a peer-reviewed publication of commentary on technology and culture; and of 40,000 people collaborating to create a more efficient human-edited directory for the Web than Yahoo. He discusses the online, peer-produced encyclopedia, *Wikipedia*, which some argue is more reliable and is certainly more current than the *Encyclopedia Britannica*. He writes about the gaming culture, where thousands of players also function as developers of their ongoing and ever-changing virtual worlds.<sup>17</sup>

Benkler examines in detail the economic, social psychology, legal, and industrial engineering principles that undergird this new phenomenon.<sup>18</sup> Some of these issues contribute to a vision for an “archives by the people.” Benkler discusses the culture—the mores and social norms—that grow, often from the community in very democratic ways, for regulating participation and behaviors. He considers how respect for competencies regulates behaviors. In short, his case studies demonstrate that peer-production does work, at least as it applies to information and culture.<sup>19</sup> The key to understanding why it works is to realize that information is a “nonrival” commodity,<sup>20</sup> that is, “its consumption by one person does not diminish its availability for use by any other person,”<sup>21</sup> or, as Shalini Venturelli has argued, unlike other commodities, “information products are not consumed one unit at a time. Rather, each product unit is designed to be utilized by many, thus becoming more valuable with use.”<sup>22</sup> An archival record bears out this conclusion; it is a nonrival commodity that becomes more valuable the more people use it.

Each of Benkler’s examples of peer-produced works operates in a slightly different way, but similar principles and factors guide them all. One is the

<sup>17</sup> Benkler, “Coase’s Penguin,” 386–90.

<sup>18</sup> Benkler, “Coase’s Penguin,” 371–81.

<sup>19</sup> Benkler, “Coase’s Penguin,” 436–41.

<sup>20</sup> Clearly, proprietary information can be a valuable asset because it provides an advantage over rivals while it is controlled and hidden. I assume, however, that archives accession records for the express purpose of making them accessible and therefore are intended to be “nonrival.”

<sup>21</sup> “It has been commonplace for a long time to treat information as a perfectly nonrival good.” Kenneth J. Arrow, “Economic Welfare and the Allocation of Resources for Invention,” from Part VI, Welfare Economics and Inventive Activity, in *The Rate and Direction of Inventive Activity: Economic and Social Factors* (Princeton, N.J., Princeton University Press, 1962), 616–17.

<sup>22</sup> Shalini Venturelli, *From the Information Economy to the Creative Economy: Moving Culture to the Center of International Public Policy* (Washington, D.C.: Center for Arts and Culture, 2001), 7–8. See also <http://www.culturalpolicy.org/pdf/venturelli.pdf>, accessed 20 March 2006.

need for a sponsoring agent who often also acts as the aggregator of the peer-made product.<sup>23</sup> Benkler sees the networked digital environment as an opportunity for intelligent people to satisfy their curiosity and contribute to society.<sup>24</sup>

Similarly, this model portends an archival system that uses the eyeballs and the intellect of thousands of volunteers—including archival customers, historians, genealogists, students, and others—throughout the world. Acting as partners with archivists, users can do what archivists alone cannot do. Archives do not have the resources to do item-level description and indexing. But archivists can become organizing agents for others to do such work,<sup>25</sup> either independently or as part of social tagging projects.<sup>26</sup>

The model calls for each archival institution to manage the aggregation and delivery of the data collected. The system might ingest the contributions of the volunteers, replacing, perhaps, the minimum metadata, extended now to something more complete. Or the contributions could become a blog of comments, supplementing the metadata. “Web-based annotations are a means by which group members create and share commentary about documents.”<sup>27</sup> Naturally, all of these data would be fully searchable, enhancing the provenance method for access. Indeed, the aggregating function may bring together data describing collections and parts of collections within the repository and, perhaps, across repositories.

Such an engineered archival system focuses the work of professional archivists on doing what they do best and what the lack of resources requires of them: organizing and describing their holdings as aggregates, not as discrete items. In addition, this new order requires that they also organize processes that invite participation in the archival commons, shared mutually by archivists and by archival users.

<sup>23</sup> Benkler, “Coase’s Penguin,” 406.

<sup>24</sup> Benkler, “Coase’s Penguin,” 426–32.

<sup>25</sup> The Greene and Meissner study implies a need to rethink divisions of labor and changing roles in archives. So does my suggestion that archivists might become, among other things, managers of digital volunteer efforts.

<sup>26</sup> A proposal for creating digital libraries, and not just indexing their content, is found in Aaron Krowne, “Building a Digital Library the Commons-based Peer Production Way,” *D-Lib Magazine* 9, no. 10 (October 2003), available at <http://www.dlib.org/dlib/october03/krowne/10krowne.html>, accessed 24 August 2006. An archives-related research project is reported in Magia Ghetu Krause and Elizabeth Yakel, “Interaction in Virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid,” *American Archivist* 70 (Fall/Winter 2007): 282–314.

<sup>27</sup> This idea is consistent with Amazon.com’s method of soliciting readers’ comments about books. There are a number of other examples on the Web. Its application for archives was introduced in a 2002 article by Michelle Light and Tom Hyry: “Colophons and Annotations: New Directions for the Finding Aid,” *American Archivist* 65 (Fall/Winter 2002): 226–29, which is the source of this quotation.

### **Why Would Anyone Participate?**

Benkler reflects on the issue of motivation and puts forth several theories about why people volunteer for this kind of work. He proposes three incentives: monetary rewards, intrinsic hedonic rewards, and social-psychological rewards. The extent to which money is a direct factor determines whether the other two are important. Money, as an indirect factor, however, can play a role. Reputation and status can be turned to consulting, speaking, books, promotions, pay raises, and other types of money-making endeavors. However, most potential volunteers who have discretionary time can choose whether to watch television or be engaged in intellectually stimulating activities or socially important undertakings.<sup>28</sup> Not so hard to imagine; just think of the army of genealogists building family trees over the Internet through name indexing and vital records data collection projects.<sup>29</sup>

The ability of the volunteer to choose projects to engage in and to determine how much time to give distinguishes peer-production in a networked environment from other voluntary work. In this microlabor market, each individual chooses his or her job, selects his or her own hours, and earns his or her own psychic reward. The development of Web 2.0 tools and participation in such collaborative efforts as del.icio.us and Flickr demonstrate the truth of these assumptions. The growing phenomena of folksonomy and social tagging demonstrate that interested individuals will devote their time and energy to make sense of the World Wide Web.<sup>30</sup> The archivist's job is to make sure that this tagging supports archival access systems.

Archivists have among their customers a natural pool of volunteers. These customers have strong incentives to become suppliers of detailed data about

<sup>28</sup> Benkler, "Coase's Penguin," 426–32.

<sup>29</sup> GenWeb is an example of a loosely organized group of genealogists who independently select vital records, index them, and publish the indexes online. There is no quality control, true aggregation, or direct linking to the original. What you see is what you get: many separate online indexes, loosely organized under the USGenweb Project umbrella, available at <http://www.usgenweb.org>, accessed 20 March 2006. The Ellis Island project is an example of a more organized effort to index genealogical records, available at [http://www.ellisisland.org/genealogy/ellis\\_island\\_search\\_tips.asp](http://www.ellisisland.org/genealogy/ellis_island_search_tips.asp), accessed 7 June 2007.

<sup>30</sup> See Adam Mathes, "Folksonomies—Cooperative Classification and Communication Through Shared Metadata," produced at the Graduate School of Library and Information Science, University of Illinois Urbana-Champaign, 2004, available at <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>, accessed 23 August 2006. The term *folksonomy* is, according to Thomas Vander Wal, who coined it, "the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval. The tagging is done in a social environment (shared and open to others). The act of tagging is done by the person consuming the information. The value in this external tagging is derived from people using their own vocabulary and adding explicit meaning, which may come from inferred understanding of the information/object . . . . The people are not so much categorizing as providing a means to connect items and to provide their meaning in their own understanding," available at <http://www.vanderwal.net/random/entrysel.php?blog=1750>, accessed 23 August 2006.

archival holdings. Benkler's work demonstrates that if one volunteer indexes records she or he is interested in, that volunteer contributes to the social good and produces personal gain. More important, difficult problems can benefit from multiple indexers and reviewers.

The cost for each person, in the first case significant, becomes trivial in the second. One person working for a thousand hours usually requires a promise of monetary reward to justify devoting that much time. Since indexes can not be copyrighted and do not have large markets, those rewards are limited. Price signals suggest this may not be a lucrative activity unless one limits access to the added-value work. But, controlling access through limited sales or prohibitive pricing restricts the social good that could come from the project.

The second case illustrates why it should work, especially for the genealogical community, which has a tradition of sharing its research.<sup>31</sup> The cost for each of a thousand volunteers is small. Each volunteer as information user enjoys the benefits, often more quickly.<sup>32</sup> The "ownership" is distributed among the contributors who in effect give it away freely to the public by having the archives publish online.

However, unless these projects are carefully managed, the advantages of a peer-based system can become disincentives. Peter Hirtle, in his 2003 SAA presidential address, "Archives or Assets?",<sup>33</sup> points out that the major "asset" an archives has is the value added by archivists in the description processes. Arguably, an institution could charge for such value-added information.<sup>34</sup> But, if a community of interested users adds the value instead, those volunteers, not the archives, are the collective owners. They will resent attempts to commercialize their work, especially if asked to pay for access to what they freely contributed in good faith with the expectation that all will benefit. Preventing

<sup>31</sup> The Genealogical Society of Utah (GSU), as part of its effort to collect microfilm (and more recently, digitized) copies of vital records, developed an Internet Indexing System to manage work flow and provide robust tools for volunteers to carry out name indexing projects from the comfort of their homes. The resulting data goes not only to the GSU's cumulative database, but to each repository that owns the records for merging with its discovery system. This system varies from the work of the GenWeb in this important way. The Georgia Archives and the Ohio Historical Society have participated in successful tests of this system. For more information, see <http://www.familysearchindexing.org>, accessed 24 September 2007. A scholarly study is Elizabeth Yakel and Deborah A. Torres, "Genealogists as 'Community of Records,'" *American Archivist* 70 (Spring/Summer 2007): 93–113.

<sup>32</sup> For example, a scholar researching a collection of historical records could organize (using students or other scholars) a project to transcribe the records to make them easier to use.

<sup>33</sup> *American Archivist* 66 (Fall/Winter 2003):235–47.

<sup>34</sup> A commercial example of this is Ancestry.com, which digitized and thoroughly indexed some of the U.S. Federal Census returns, adding value to a public record. The records themselves (or a microfilm thereof) are available for all to see at no cost at the National Archives and at many other repositories. But Ancestry charges for access to its database and with it, its digital images taken from the microfilm. The added values, those that make this worth paying for, are 1) the aggregation, 2) the indexes, and 3) the convenience of online access.

third-party commercial appropriation of the aggregated work may require the controlling agents—each archives or its agent—to manage rights, using something similar to the General Public License pioneered by the open-source software community.<sup>35</sup>

As public trusts, the nation's archives must continue to be open and accessible to all without cost. More than a mere statement of principle, this is a necessary element of the peer-production system. The results: archives whose holdings are much easier to discover, access, and use. And the bonus is a community of highly intelligent men and women who will come to understand and appreciate archives. The archives of the people (as they have always been, but only in the abstract) thus become the archives by the people (who contribute and add value) and for the people (who now can actually use them). Archives thus become a part not only of the information economy, but of the knowledge and creative economy. A “networked information system levitates the value of ideas and forms of expression [and] causes further heightening of demand for the same expression, . . . creating an upward spiral.”<sup>36</sup> Archival institutions must reinvent themselves, in collaboration with other archives and with other types of organizations, to systematically invite and encourage commons-based peer-production. The archives of, by, and for the people demands no less.

<sup>35</sup> Benkler, “Coase’s Penguin,” 379. Find more on the General Public License at <http://www.gnu.org/copyleft/gpl.html>, accessed 2 March 2006.

<sup>36</sup> Venturelli, *From the Information Economy to the Creative Economy*, 8–9.