

The Invisible Hand and the Accidental Archives
Max J. Evans, Executive Director
National Historical Publications and Records Commission

Choices and Challenges Symposium
The Henry Ford
October 8, 2004

“How are you doing?” I asked a friend at the AASLH Annual Meeting. Her answer “If I’m surviv’n, I’m thriv’n!” I like the attitude. But I think the issue posed by this session “Thriving in the Digital World” envisions something more than survival.

Archives do survive. The record of the past—what is found in formal settings such as archives, libraries, and museums, or that turns up in attics, basements, and chicken coops—seems to survive in spite of society’s indifference. As institutions, archives survive, if only just barely, but do a pretty good job of carrying out their dual mission:

- Preserving and transmitting culture across generations: keeping the record of the past for the present and documenting contemporary society for generations yet unborn; and
- Serving immediate public interests by preserving records documenting personal rights and entitlements, and assuring that government and other public and private institutions maintain continuity and are accountable to the public.

But thrive? I don’t know. Thriving would mean that their commitment to public value is recognized by public support. Most don’t meet that test. Most are figuratively and sometimes literally shoved off into the backrooms of government and other organizations and treated either as a necessary, but not compelling part of the business, or as an expensive luxury. Neither view puts the archives on a footing equal to the police or fire department, public schools, or even the public library, parks, and museums. And for a very good reason, which I state here, as a working hypothesis for examination:

There is an Archives dilemma, namely that the digital age creates expectations that cannot be met using traditional archival methods. But using technology and practices adapted from a global, market-driven, connected society may offer solutions.

What are the expectations? Sophisticated, maybe jaded, researchers probably expect little more than what they get. Decent finding aids, accessible and comfortable facilities, and knowledgeable staff. Sure, they would like to have more detailed finding aids, but understand that time invested in detailed work on one collection is time lost describing many others.

And this is the nub of the problem: Faced with a choice, to invest a lot of effort on a few records, or to spread the effort thin across the whole, archives have chosen the later. Archives, perhaps uniquely among cultural institutions, are wholesalers. They provide, not finished products (historical interpretation), but better interpretation. They are like the plastic company BASF: they advertise that they “don’t make products, they make products – better.” An archives reach is extensive, but indirect. Perhaps we should borrow language from economics to describe this. Economists estimate the multiplier effect of dollars in the economy; archivists might ask what is the effect of the use of archives? Each research visit by Ken Burns means that millions of PBS

viewers learn a bit more American history, if each use by a genealogist reaches only a limited audience. Still, archivists can rightfully make the claim that they make knowledge of American history—better.

More than any other kind of cultural organizations, they contain massive volumes. The numbers are staggering. I conservatively estimate that there are at least 10 million cubic feet of records held by archives nationwide.¹ Ten million cubic feet equals at least 30 billion pages.

And it is growing. Archives, after all, operate warehouses! Among the challenges archivists face is appraisal of a glut of records created by modern bureaucracies. Archivists traditionally choose to keep only a small fraction of the mountain of paper and electronic documents. But given the emphasis on social history and the popular interest in genealogy, archivists must face more difficult decisions. Its not just macro-level records, those that document policy decisions and decisions at the highest level, but the micro-level needed to support research in the lives of everyday citizens touched by governments and other institutions.

Archivists, in the early years of the National Archives, engineered methods to take care of, or at least manage, this mountain. They deal in the aggregate: Appraisal by the series, not document-by-document; Preservation by environmental controls; Arrangement and Description, at best, at the file unit level; Reference; by delivering records by the cubic feet. There is no other choice.

No one is happy about this. Scholars would like to have item-level description. So would archivists. And a more naïve public (and sometimes administrators), expect not just item-level description, but digital images, with searchable text, of each of those 30+ billion pages!

Compare to the size of the rest of the Internet: “The Web includes some 800 million documents on computers all over the globe,” and “grows at the rate of 7.3 million new pages a day.”² Sound impressive? Well maybe not, when you consider that at that rate 11 years will pass before the Internet will have as many pages as the nation’s archives!

Given the size of the Internet, it is not unreasonable for its users to think that it contains everything. And for quick and easy reference, it’s pretty handy. “If we don’t have it, you don’t need it.”³ Or, in other words, if it’s important, it will be there. Does this lead to unreasonable expectations that cannot be met? I think so.

What is needed, instead, is a strategic approach to *selective* publication of archival holdings. By *publishing* I mean getting the content of primary sources into the hands of people outside of repositories. Strictly speaking, I suppose one could argue that making archives accessible in the research rooms is to publish, but that defines a commonsense understanding of the word. People think things are “published” when they are easily accessible and usable.

¹ Based on the assumption that the state archives and local government and all other archives each equal or exceed the volume of records in the National Archives—3.1 million cubic feet

² Source:http://www.thefab.net/topics/computing_general/cg21_how_big_is_internet.htm

³ Advertising slogan for Allied Development, a Salt Lake City retail operation for many years that started as a war surplus supplier.

Scholarly editions—books—are one variety of this, but, digital content, including images of primary sources published online on the Internet, qualifies. The later is, in my opinion, the future. However, only documents important enough to justify the costs should be published. This gets to the selectivity. How to decide? What do we mean by “important?”

The scholarly community has created an unofficial canon: the papers and records of the founding fathers and a few dozen other collections that illustrate American history topics selected for full scholarly treatment. Somewhere between these works, found between hard covers, and the huge mass of materials found only in grey archives boxes and relatively untouched, are other documents that could be “published” as digital images. Because what is important to a scholar may not be important, say, to a woman looking for her grandfather’s will or man looking for photographs of historical railroads. Documenting democracy demands a documentary democracy that treats all people fairly and equally.

Archives do, of course, selectively publish their holdings online, identifying what they estimate will be most useful for research. Or they make digital exhibits of what tell compelling stories. These digital exhibits are often funded by grants. And there are many wonderful web sites that use digital facsimiles to help tell America’s stories.

But neither printed nor digital editions are solidly incorporated into the fabric of archival practices. And the selection processes take place in the rarified atmosphere of humanities scholarship. Nothing wrong with that, as far as it goes. But we need additional models, *ones that rely on the individual choices of the people*.

One model to consider is Amazon. Not the race of war-like women from mythology, or the South American river, but the dot-com. How does it work?

1. No books.

How did Amazon become the world’s largest bookseller, but not have any books? Well, it’s not true they don’t have books, they do, but they are found in a sophisticated just-in-time purchasing and delivering system. They are not found on shelves for customers to browse. In this way, they are like archives: they have the records, but don’t allow browsing.

2. Rich graphics interface.

Instead of browsing the shelves, Amazon uses a rich graphics interface that allows searching and browsing. It is not boring. A comparison to archives reveals opportunities for improvements.

3. Uses metadata in interesting ways.

Including using images as metadata. See above comment and the one below.

4. Employs varying levels of description.

Some books are more equal than others. Meaning that some are listed with minimum bibliographic treatment, others include tables of contents and/or indexes and/or selected pages. Other include the full text of the books. Archivists have learned that lesson. They may arrange and describe at the series, subseries, file unit, or document item level.

5. Seeks implicit feedback from customers.

Amazon customers can read what other customers think about the books, but just what the publisher puts on the dust jacket. These reviewing customers become part of an anonymous community of participants in decision-making. Archivists should develop systems that invite participation.

6. Smart system.

Amazon keeps track of the behavior of its customers and responds accordingly by providing suggestions for additional reading. Archivists should invent systems that map researchers' questions through the maze of archival principles, practices, finding aids, and informal knowledge of working archivists, to the records needed.

How can Amazon instruct archives? The rest of my paper seeks to answer that question with a demonstration of a virtual archives. That is, it contains what I hope you will find to be interesting and useful *representation* of archival holdings.

Two caveats:

1. This is *not* a real system. It exists only in my wild imaginings, and is subject to change in its detail without notice. Furthermore, it is a virtual representation of an imaginary archives.⁴
2. I'm not a web designer. The illustrations are to make points, and have not been professionally designed. The point is not in the details, but only in the concepts they illustrate.

This virtual archives intends to illustrate:

1. Graphic interface.
2. Variable descriptive levels.
3. Market-driven selection processes.
4. That digitization can be affordable.
5. How to invite participation in adding value.

The HistoryMax Virtual Archives

I've indicated that archives are like wholesalers. Or maybe they are like banks; stable, secure, reliable. (and seemingly inaccessible, hard to get from them what you want, and keeping "bankers' hours.") How can archivists send the positive message to a general public? The Internet proves the old saw that "a picture is worth a thousand words." So let's try graphic representations. A user interface may start with a symbol of security: the vault.

There are no guards, and the door's open. Let's go inside.

The graphics illustrated here point out how users might envision an archives scope by browsing. However, the virtual archives must also be searchable. It should have in addition, like Amazon, intelligence to help interpret naïve search requests and provide alternatives paths, and to remember and remind users of previous searchers, for example.



⁴ However, the collection used for illustration, the Susa Young Gates Papers, is from my former institution, the Utah State Historical Society. I thank my former colleagues for permission to use images from the collection. Mrs. Gates was Brigham Young's daughter, a leader among Mormon women, and an ardent suffragist.

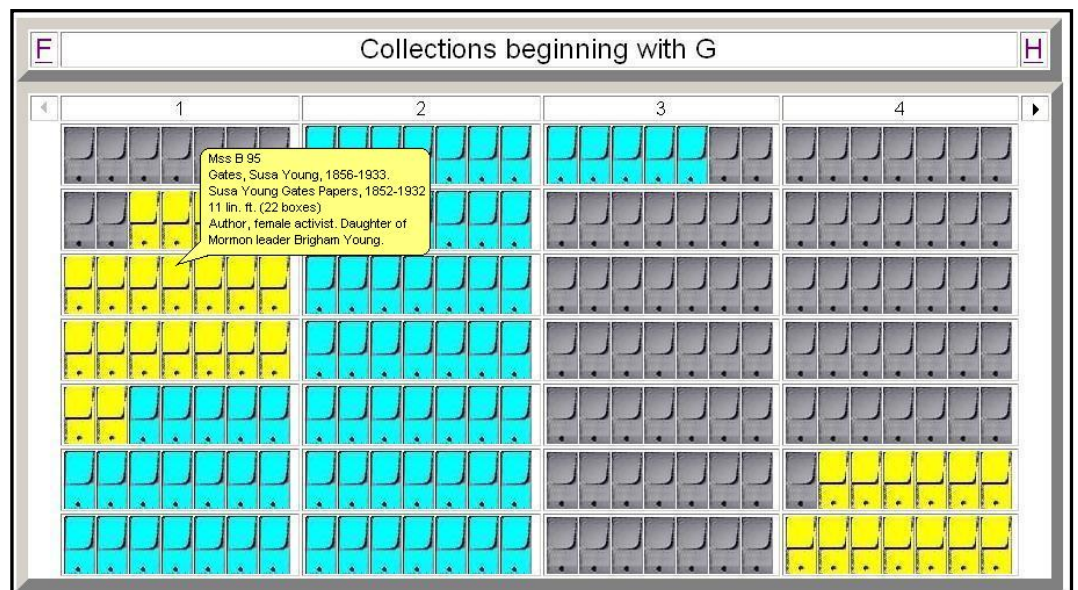
Click on the vault for a virtual browse. The vaults—the stacks—are filled with row after row of archives boxes, organized, as shown here, alphabetically by collection title. Since this is a virtual archives, however, they could be displayed in many other different ways: by record group, by geographic coverage, by inclusive dates, or by any other logical and useful principle. A virtual archives allows alternate views.



Researchers browsing the virtual stacks can choose a section to delve deeper. Clicking *Collections beginning with G*, for example, takes you a representation of the holdings as found in archives boxes. Box colors change at the beginning of a new collection. The number of boxes represents the relative size of each collection.

You may *point to a collection* to learn more about it by reading a brief abstract in the balloon.

This illustrates how collection-level data, from the MARC catalog or from a collection of EAD files, can graphically represent the collections in the aggregate. It shows how users could experience browsing the stacks, getting a sense of the contents of collections, in a way more visual than scanning lists.



Engaging Users

Many collections are unprocessed, but, perhaps minimally cataloged so researchers can know of them, *and, so the archives can engage the researchers in processing decisions.*

George G. Gandy Papers (Mss B 1989)
This collection is not yet available for research.
 Please complete this form to request archival processing.

Collection	George G. Gandy Papers
Collection ID	Mss B 1989
Size	9 boxes
<input type="radio"/> Please process the above named collection. Notify me by email when the collection is completed and ready to use.	
<input type="radio"/> Please process the above named collection. I expect to be in your repository on <input type="text"/>	
Name	<input type="text"/>
Address	<input type="text"/>
City, State, Zip	<input type="text"/> <input type="text"/> <input type="text"/>
Phone number	<input type="text"/>
Email	<input type="text"/>

A user *selects an unprocessed collection* from the shelves. The next screen is not a detailed finding aid, but a form with minimum description, from the MARC catalog or EAD record, giving the user the opportunity to request processing.

This example offers two choices: (1) a soft request to process the collection and (2) a notification of intent to visit.

Most archivists object to the later, what I call *just-in-time processing*. However, both offer

opportunities to engage customers, in ways not unlike Amazon. Archivists have choices about what to process next; why not engage users in the decision by direct methods like this? What could be more democratic and responsive?

If you go back to “*Collections beginning with G,*” and click on the Susa Young Gates Papers, you may see a traditional archival finding aid, complete with a container list, describing a processed collection.

Not shown here is Summary Description, Background, Scope and Content, and Administrative Info, or the ten pages or so of container list. All of that is to be found online, as it should be, for those who have the patience to read all this text.

However, keeping with the graphics theme, I suggest that there may be other ways to represent this information. You could get a graphic, virtual view of the contents of this collection.

The Susa Young Gates Papers, 1852-1932

A Register of the Collection at the Utah State Historical Society

The machine-readable finding aid for this collection was created by the Collections Management staff, Utah State Historical Society, with financial assistance from an LSTA grant provided by the Utah State Library Division.

Utah State Historical Society
Salt Lake City, Utah

Copyright, 2000, Utah State Historical Society. All rights reserved.
 Reproduction, storage or transmittal of this work, or any part of it, in any form or by any means, for commercial purposes, is prohibited without prior authorization of the Utah State Historical Society. This work may be used for scholarly and other non-commercial use provided that the Utah State Historical Society is acknowledged as the creator and copyright holder.

[Summary Description](#) | [Background](#) | [Scope and Content](#) | [Administrative Info](#) | [Container List](#)

Container list

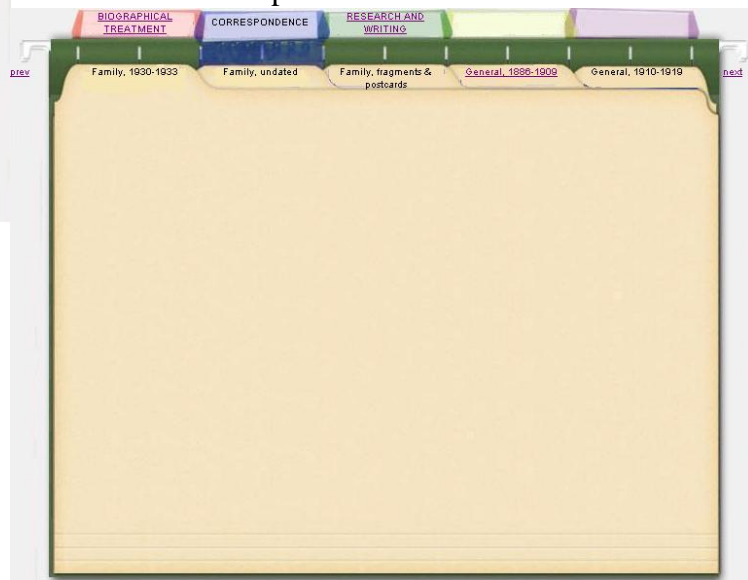
Box	Folder	Contents
Biographical Treatment		
1	1	Biography <ol style="list-style-type: none"> 1. Excerpts, "Susa Young Gates: Her Life and Times," Thesis by Paul Cracroft, 1951 2. Tribute to Susa Young Gates, <i>The Relief Society Magazine</i>, July 1933
1	2	Autobiographical notes, ca. 1928; Transcript of diary, 1870
1	3	Biography <ol style="list-style-type: none"> 1. Sketch of Susa Young Gates, by John A. Widtsoe 2. Sketch of Mrs. Leah E. D. Widtsoe 3. Notes on church experiences, testimony, priesthood
1	4	Bibliography of periodical articles by SYG
1	5	Family materials <ol style="list-style-type: none"> 1. Notes of family memories 2. Lion House 3. SYG on death 4. Fragments on Brigham Young in Nauvoo 5. Brigham Young's will 6. Blueprint of ground floor, Lion House, 1932
1	6	Life Story of Lucy Bigelow Young
Correspondence		
Family Correspondence		
2	1	1852-1889
2	2	1890-1919
2	3	1920-1927

This graphic shows that the collection is subdivided into series, represented by the top tabs. The file units in each series are represented by file folders. Data in the labels for each comes from the same EAD file that produced the textual finding aid.



The Biographic Treatment series, with six file units, includes four that have item-level descriptions. The one shown, named Family Matters, lists six items.

A view of the Correspondence series illustrates another point.



Each folder is a link, like the one labeled “General, 1886-1909.” These links could lead to one of two displays: 1) if the file has been digitized, to a virtual folder containing images of all the pages or 2) if not, to an Order Form for the user to complete.

The top of the order form is populated from the EAD data. The user completes the rest of it to request digitization, when convenient for the archives, or to place an order, much like a microfilm order or photocopy order, and pay for the cost of scanning.

Then once scanned, each time a researcher clicks on the folder tab they open a virtual folder.

Manuscript Collection Scanning Order Form	
<i>The file described below has not yet been scanned. Please complete this form to request scanning.</i>	
Collection	Susa Young Gates Collection
Series	Correspondence
Subseries	General Correspondence
File	1886-1909 (Box 3, Folder 1)
<input type="radio"/> Please scan the above named file and charge my credit card at the rate of \$0.xx an image. Notify me by email when the file is completed and ready to use on the Internet.	
<input type="radio"/> Please scan the above file at your convenience. Notify me by email when the file is completed and ready to use on the Internet.	
Name	<input type="text"/>
Address	<input type="text"/>
City, State, Zip	<input type="text"/> <input type="text"/> <input type="text"/>
Phone number	<input type="text"/>
Email	<input type="text"/>
Credit Card	<input type="text"/>
Etc.	<input type="text"/>

The virtual folder opens at the first page of the first document in the file. The left-hand list of documents and pages allows for virtual “thumbing” through the file. The headings in this list are simply document numbers, not descriptions or titles.

Using Existing Metadata

This illustrates my next point, that one of the reasons digitizing whole collections, or significant parts thereof, is expensive is because of the high cost of providing detailed metadata for each item. Current digitization standards and best practices, developed and used by the digital library

community, go beyond simple item-level description. They are based on the assumption that collections are aggregates of individual items. But archival methodology assumes bodies of materials are made up of organically-related items. One method starts with the trees, the other with the forest.

Archivists can employ archival methods to digitize archival units and do it at a reasonable cost, by importing existing metadata, extracted from EAD files.

Susie Young Gates Collection Correspondence General Correspondence 1886-1909 (Box 3, Folder 1)	
Document #001 P.001 P.002	Collection: Utah State Historical Society - Susie Young Gates Collection
Document #002 P.001 P.002 P.003	Publisher: Utah State Historical Society Identifier: Ms B95
Document #003 P.001 P.002 P.003 P.004	Collection: Susie Young Gates Collection Series: Correspondence Subseries: General Correspondence File Unit: 1886-1909 Container: Box 3, Folder 1 Document Title: Document #001 Page ID: P.001
Document #004 P.001 P.002	Author: Date: Subject:
Document #005 P.001	Person:
Document #006 P.001 P.002	Corporate name: Geopolitical place: Description: Comment:
Document #007 P.001	Rights Management: Digital Image (c) 2001 Utah State Historical Society. All Rights Reserved.
Document #008 P.001 P.002	Holding/Institution: Utah State Historical Society Source: Ms B 95, Susie Young Gates Collection Source item number: B0095003001001001
Document #009 P.001 P.002	Source format: Source size: Source physical description: Source donors: Source donation date: Type: Format Use: image/jpeg Format Creation: Resolution: TIFF: 300 ppi Bit depth: 8-bit grayscale Dimensions: JPEG: 700 x 556 pixels Scanning device: Creo-Scitex Jazz+ Flatbed Scanner Date Digital: 12/11/2001 Scanned by: Utah State Historical Society, Haley Peterson Metadata cataloger: Haley Peterson Collection information: http://history.utah.gov/FindAids/E00095/
Return to Finding Aid Show list	

However, they must come to accept that it is reasonable, if not ideal, to present digital images without detailed metadata, if maintained in context, and browsable.

The table of metadata associated with the image of this page illustrates a minimum level approach to metadata.

I imagine an archival operation that takes requests for scanning as routinely as it does for photocopying. The form shown earlier is used as part of this process. The processes of photocopying and scanning are similar: its page-by-page handling and picture-taking. While scanning, the operator would use a tool to open the EAD file for this collection and

find the file unit being scanned. The tool would extract all of the data, including the higher level data, from the EAD. It would also use global settings for the technical data, and populate this table without much user intervention. The operator would only change the page number after each scan, and the document number after each document. She or he would not add author, title, subjects, and other potentially useful access points. The results, as shown above, provides browseable use of the folder at remote locations.

It is no worse than microfilm, or for that matter, browsing the hard copy in the research room. And in many ways it is much better, because it allows for use of the materials outside the repository. Microfilm does do, but this is better because it is possible to produce better images (in full color) and to navigate without cranking through thousands of feet of film.

If it is not perfect, then I reminded you that the perfect is the enemy of the good. Furthermore, it establishes a platform to launch another phase: volunteer data extractor. *Click on the Volunteer Extraction Form* link.

The top of this form is populated from data in the metadata. A volunteer completes the rest of it. Voluntary efforts might be random and idiosyncratic, or organized and controlled, depending on policy set by the archives.

The data is submitted and entered, updating the metadata.

Volunteer Document Extraction Form

The Archives encourages volunteers to transcribe or otherwise extract information from pages published on our web site.

If you want to volunteer, please complete this form with as much information as you can.

Collection	Susa Young Gates Collection
Series	Correspondence
Subseries	General Correspondence
File	1910-1919 (Box 3, Folder 2)
Document	1
Page	1
[The data above comes from the data in the EAD finding aid.]	

PLEASE PROVIDE YOUR CONTACT INFORMATION

The archives will not publish your contact information, but use it only to reach you with questions about the extraction. It will, with your permission, credit you as the contributor.

I want my name included with the extracted data.
 I do not want my name included with the extract.

Name	Jeanne O. LeJest		
Address	6 Family Tree Circle		
City, State, Zip	Researchville	FH	99876-6547
Phone number	234-555-6543		
Email	jeanneo@myfamily.org		

USE THIS FORM TO COMPLETE EXTRACTION

Enter Summary Information

[Letter] [June 23, 1898] [June 23, 1898]

Document title/written subject (if any) [Letter from John R. Park to Editor, Deseret News]

Physical description [First page of two page, typewritten letter, 8-1/2x11]

From/To [Park, John R.] [Editor, Deseret News]

If you provide a transcript, indicate the type:
 Full
 Abstract/Summary
 Transcript

Use the forms below to enter index terms.

Use the enter key to separate one entry from the next.

Enter names of people, including families (last name, given names) Park, John R.	Enter place names Salt Lake City, Utah	Enter organization names Deseret News	Enter topical subjects Mormonism Psychology
---	---	--	---

The data elements shown in red on this updated page description are those entered by the volunteer. Note that the Document #001 is replaced in the frame on the left with the newly-supplied title.

Susie Young Gates Collection Correspondence General Correspondence 1886-1909 (Box 3, Folder 1)	
Letter from John R. Park to Editor, Deseret News P.001 P.002	Collection: Utah State Historical Society - Susie Young Gates Collection Publisher: Utah State Historical Society Identifier: Mss B95
Document #002 P.001 P.002 P.003	Collection: Susie Young Gates Collection Series: Correspondence Subseries: General Correspondence File Unit: 1886-1909 Container: Box 3, Folder 1
Document #003 P.001 P.002 P.003 P.004	Document Title: Letter from John R. Park to Editor, Deseret News Page ID: P.001 Author: Park, John R. Date: June 23, 1886 Subject: Mormons and Mormonism Psychology
Document #004 P.001 P.002	Person: Park, John R. Corporate name: Deseret News Geopolitical place: Salt Lake City, Utah
Document #005 P.001	Description: First page of two page, typewritten letter, 8-1/2 x 11 Comment:
Document #006 P.001 P.002	Rights Management: Digital Image (c) 2001 Utah State Historical Society. All Rights Reserved. Holding Institution: Utah State Historical Society
Document #007 P.001	Source: Mss B 95; Susie Young Gates Collection Source item number: B0095003001001001 Source format:
Document #008 P.001 P.002	Source size: Source physical description:

It's about engaging archives users. It makes them part of a corps of volunteers whose self-interests merges with those of the archives, to produce an ever-growing body of richly-described and indexed collections. It's Amazon-like in this way. It takes advantage of organized, or self-selected and anonymous users who can work at home and in remote locations.

Conclusion:

In *Wealth of Nations*, Adam Smith wrote that individuals "...[B]y directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was not part of it."

As scholars—and others—use and interact with the archives as I've described, they have opportunities to enrich it, as if by accident, but these accidental archives are really engineered archives, built by design and choice that makes end users our partners in preservation and access.

Necessity is the mother of invention. Clearly archives will never digitize all in their holding—nor should they. But while they continue to make the original records accessible in person, they should publish catalogs and other finding aids as means to engage the public in archival decision-making and in voluntary service.

The end will be more historical records accessible and, therefore, more readily used by more people in more ways. And more people engaged and committed to this important work. Indeed, these means may deliver in the end the promise of a documentary democracy, where people can easily, if not instantly, get what they need for their own use.

That's what I call creating public value!