

Tamiment Web Archiving Collecting Policy

Updated 9/8/2010

Introduction

A summary of Tamiment's web archiving activity during the NDIIP grant era reveals that Tamiment's core web collecting mandate, i.e., archiving the websites of entities with whom we have a collecting relationship (e.g. NYC labor unions), has been fulfilled, along with having largely archived the web presence of the U.S. Marxian left., and that significant initial progress in archiving the web activity of the alternative mass media and contemporary anarchism has occurred, and that the Library created the Guantanamo Bay Prison Camp Web Archive to complement Tamiment's archival collecting initiative in this area. All told, Tamiment has created 18 web archives (10 publicly accessible), and archived 526 websites, whose 1589 captures to date (8/03/2010) occupy 1.6 terabytes of storage space.

Collecting Scope:

The Tamiment Library and Robert F. Wagner Labor Archives at New York University form a unique, internationally known center for the study of the history of Labor and Radicalism, as well as the social and cultural contexts in which these movements function. Part of Tamiment's mission has always been to document the political communication of the radical and labor entities whose records we preserve. Traditionally this communication has taken place using flyers, pamphlets, handbills, broadsides, and brochures. Today much of it takes place on the web.

In order to assure that the Library continues to document the changing nature of U.S. labor and radicalism, the Library has come to see web archiving as an essential component, and not only for documenting the migration of the activity of long-standing radical entities and related topical/functional activity to the web, but to documenting new issues, ideas, forms, and content of labor and especially radical activity, whose emergence/leading edge is perhaps best reflected on the web.

In principle, the collection development policy for the Tamiment Library's Web Archive parallels that of the Tamiment Library as a whole (labor and radicalism), as well as the social and cultural contexts in which these movements function), including in regard to geographical¹ and chronological² scope, and collecting levels.

In practice, this is complicated by (a) the enormous size and variety of the universe of born digital materials within Tamiment's collecting scope, (b) the emergence of numerous and diverse new areas of progressive activity, thought and culture, and of born digital documentation created thereby (or there about), and (c) resource restraints. Thus the Library will not only have to carefully appraise materials, but to set priorities and limitations.

¹ Principal areas of concentration are national and local level activity, while also documenting the international context and connections that increasingly part of U.S. labor and radicalism, and selectively documenting activity in other portions of the U.S.

² Naturally, given the recent development of the internet/www, the our web archive will be weighted towards more recent and contemporary documents, although we will also collect websites that document the history of labor and the left, and which will contain more older material.

Crucial to this endeavor is a clear understanding of what constitutes radicalism (a multi-faceted concept), and the consistent (although not inflexible) application of that understanding to the Library's web archiving activity, such that each web archive is a coherent and (near) comprehensive representation of radical activity within its given area.

Radicalism, as implied by its etymology, is first and foremost a fundamental rejection of the foundational political/economic principles (and their associated social and cultural manifestations) of the contemporary U.S. or world economic order, i.e., capitalism.

Another meaning of radicalism refers to the use of militant, direct action tactics, ranging along the spectrum from civil disobedience all the way to terrorism, including such intermediate measures as occupation of institutions and/or property, and the smaller-scale creation of alternative social arrangements.

In addition, radicalism includes not only challenges to existing social relations in toto, but also, (where the challenge/critique, especially in its given historical context, is more fundamental and militant) to important components thereof, e.g., feminism, gay rights, anti-racism, opposition to car crimes and to limitations on civil and human rights post-911, etc. As these latter examples illustrate, it is important to consider radicalism not only as a noun, but as an adjective.

Thus, there will be certain web archives that of necessity, and appropriately, may include a significant amount of, or sometimes even a preponderance of non-radical entities (e.g., "Guantanamo/U.S. War Crimes"). Tamiment's web archiving collecting policy includes the obligation to identify -- and within the limits of our resources, collect -- the most significant topics/potential collections of this nature, based on their intrinsic worth, their relation to our overall collecting policy and collection strengths, and their relationship and proximity to U.S. labor and radicalism. (A striking example would be the contemporary manifestations of a venerable historical phenomenon, namely the effort to transform the Democratic Party into a social democratic organization.)

Traditionally, radicalism has been associated with the concept of social classes, the paradigmatic example being Marxian socialism and communism, and to a lesser extent, anarchism. Although the latter's focus has historically been on freely-associating individuals as the foundation of a just society -- an important factor in anarchism's remarkable resurgence since the fall of (statist) Communism -- also influential are the efforts of a new generation of anarchist intellectuals, who, in the search for viable real-world anarchist models, seek to transcend traditional anarchism's methodological individualism-essentialism, while retaining anarchism's anti-statist perspective.

Meanwhile, ironically, perhaps, radicalism associated with the concept of social classes has been a "victim of its own (partial) success," i.e., the absorption into and modification by the liberal capitalist order, of much of its intellectual content and proposed social reforms, exemplified by the "welfare state."

Realizing the difference between identity politics and radicalism, we have made a conscious decision not to collect in the former area, although areas such as the immigrant rights movement (or rather, principally the radical components thereof) do fall within our collecting scope.

Today, newer movements, notably the above-described “new anarchism,” the anti-globalization movement, the environmental/green movements, and feminism have become increasingly important and the source of newer and radically different overarching worldviews, even while incorporating anti-capitalist perspectives and portions of Marxian and classical anarchist thought.

Whatever the scope of an individual web archive is, selection of websites for that archive must be guided by the above-described understandings of radicalism, so that any given web archive is coherent and approaches comprehensiveness within its given scope, and of manageable size. If a web archive suffers from “bracket creep,” i.e., ill-considered expansion via the inclusion of numerous entities animated by non-radical world views (e.g. progressivism or liberalism), the archive will thereby lose its coherence, comprehensiveness, and manageability. Although as mentioned above, there are archives where significant inclusion of, or even predominantly composed of non-radical entities is appropriate (“Guantanamo / U.S. War Crimes”), such “bracket creep” could also bring Tamiment’s web archiving program as a whole up against the limits of the resource constraints within which it operates.

Resource Constraints:

Resource constraints include: (a) the cost of storage space, (b) the amount of storage space needed/available, although ongoing technological advances will alleviate these concerns. (Note that the CDL is on the verge of introducing a solution to the de-duplication problem, i.e., the recapture of unchanged web pages with successive captures of websites, which will reduce storage requirements by approximately 50%); (c) staff resources, both for collection maintenance (metadata, analysis of the quality of captures and communication with CDL regarding such technical problems, proposed enhancements, etc.), but most importantly, ensuring adequate time for the project administrator, who is primarily responsible for collection development, to search for and select carefully from the enormous number of websites that document the vast topical and functional scope and diversity of labor and radical activity occurring/represented on the web.

Currently, staff time is envisioned as follows: a minimum average of five hours/week by the Project Administrator (collection development, plus supervision and research); five hours each week by an Archival Assistant (evaluating website captures, updating finding aids, etc.), ideally an individual enrolled in either NYU’s archival management/public history program and/or the NYU-Palmer School’s dual master’s degree program.

Collecting Priorities:

1. Tamiment’s first responsibility is to continue to archive the websites of those entities with whom the Library has a collecting relationship. While its primary sense “collection relationship” refers to entities whose documentation the Library already hold, it may also be secondarily thought of as referring to entities or topics significantly represented in throughout the Library’s holdings, and to recent collecting initiatives and public programs.

2. The Library’s second responsibility is to document recent and currently emerging areas of radical activity. Foremost among these are the “new anarchism,” along with the anti-globalization movement, and the radical portions of the environmental/green and the feminist and women’s movements; radical web-based radical alternative information sources and services and web-based

radical artistic and cultural expression; while struggles around information-related issues and the cultural, political, and social content and control of cyberspace grow ever more important.

Ongoing and new areas of struggle that also merit special attention due to their significance and/or their representation in the Library's holdings and programs include: academic freedom (note the Library's Frederic Ewen Academic Freedom Center); the Guantanamo Bay Detention Camp (understood in its international context, and in connection with the Library's legal activity-focused traditional archiving initiative in this area); housing, squatters and tenant's rights; peace movements, and student and youth movements.

3. The Library's third responsibility is to document the born-digital manifestations of labor and radical activity and their cultural contexts in our traditional areas of collecting strength: the New York and U.S. labor movement; the (chiefly) Marxian left, i.e., communism, socialism, Trotskyism, etc.; classical anarchism; the so-called New Left and post-New Left; civil and human rights, primarily as they pertain to labor and radicalism; the institutionalized online manifestations of radical thought and publishing; and the web presence of notable individuals.

Of similar importance, and in keeping with the Library's focus on the interplay of labor and left politics, is to document radical electoral political parties and activity, including attempts to develop national and New York area local party or other political organizations – in particular those (typically with a social democratic orientation), that seek to create such a mass political party and/or to transform the Democratic Party into such a party (this activity is represented by a manageable number of websites).

Other Collecting Considerations:

While it would not be feasible to document even the solely radical component of the web presence of every struggle for social/economic justice and civil/human rights, the Library has already created several collections that fall under this broad rubric (and in the future will continue to create a limited number of such collections). The criteria for doing so has been, and will be, a combination of intrinsic importance, existing collection strengths, and new collecting initiatives. (Current examples include the Feminism, Guantanamo Bay Detention Camp, Peace Movements, and Prisoners Rights archives; future examples would include archives for Academic Freedom; Globalization; and Housing, Squatters, and Tenants movements.)

Although guided by an overall collecting policy, given the decentralized and diverse nature of contemporary labor and left activity, and the similar structure of the world wide web (more like a neural network than an evolutionary tree), with each website containing its multifarious links extending in all directions, the Library needs to be sensitive to the reality that unexpected collecting areas may open up, including some rather specialized ones, that call for the creation of their own web archive -- for example, the possibility that it would be fairly easy to find some two dozen websites created by radical librarians/library workers and others seeking to provide reference service, in particular to the progressive community.

New web content/formats: the Library will not confine our focus to traditional websites, but will be alert to newer forms of web activity (as they are used by labor and radicalism), such as blogs, online social networks, and if legal consensus allows, discussion lists.

Collaborative Collection Development:

While the primary responsibility for collection development should remain with the project administrator, and ultimately the Head of the Library (with important input from the Library's knowledgeable staff), the Library's overall collecting scope (labor and the left) and the size of the universe of potential documentation suggest the supplementary merits of distributed /collaborative collection development, by which is meant that suggestions for websites and even perhaps special areas of concentration worthy of leading to the creation of new topically based web archives, should be actively solicited from individuals with knowledge in depth in particular areas. (The Library's Jewish American Left web archive, which, although initiated by the project administrator, owes much of its content to the expertise and suggestions of Arieh Lebowitz, is a good example.)

NYU faculty and students, the extended Tamiment "family," and scholars and activists with whom the Library has ongoing relationships are potential sources of such information. In particular, it is important to seek out the input of the younger generation of scholars and activists, graduate students, and others with deep knowledge of the contemporary left in all its variety, and a higher degree of computer literacy and familiarity with the online world.

Collaborative collection development also means that the Library should be aware of web archives located elsewhere covering portions of its collecting scope and taking these into account in shaping its collection development policy. (A good feature of the WAS is that on the home page for each web archive is a place reserved to list related resources.)

Tamiment Library (labor and the left) web archive: Overall Usage

Storage Used	1.7 TB
Projects	19
Sites	553
Captures Completed	1688

Publicly accessible web archives ([UUhttp://webarchives.cdlib.org/institutions/NYUL](http://webarchives.cdlib.org/institutions/NYUL)) and via Tamiment's website, where each has its own finding aid whose "container lists" enumerate the websites archived under that topical area):

*Alternative Mass Media Web Archive (38) -- including radical media organizations, such as news web sites, radio, television, serials, blogs, etc., and some personal websites created for similar purposes.

*Anarchism Web Archive (19)

*Arts & Cultural Left (12)

*Communism, Socialism, Trotskyism Web Archive (59)

*Guantanamo Bay Detention Camp Web Archive (29)

*Jewish American Left (13)

*Labor Unions & Organizations (U.S.) Web Archive (141) -- NYC unions, national unions, labor movement organizations, organizations dedicated to support of the labor movement, union democracy and rank and file union groups, and labor use of the media)

*Other Left Activism Web Archive (75) -- a general category where websites are temporarily housed pending their transfer to new web archives. *In addition, this archive currently has websites that belong in existing web archives, and these will be transferred thereto when the functionality to do this automatically becomes available around the end of 2010.* (See Appendix for list of websites in this archive.)

*Peace Movements (18)

*Prisoners Rights (10) -- relates to Kunstler foundation / papers

Short term collecting priorities for the above archives:

The first priority is the expansion of the Anarchism, Guantanamo and Alternative Mass Media web archives. The second priority is to fill out the Arts, Peace, and Prisoners Rights web archives.

Collections pending public access: (40 websites)

The following (mostly representative or skeletal) collections have been created more recently, but are not yet publicly available, either due to the six month embargo between the capture and online “publishing” of any given website, and/or the fact that in some instances their size has not reached the “critical mass” deemed necessary for a publicly available web archive, often due to the above-mentioned fact that we cannot yet automatically transfer websites from our catch-all “Other Left Activism Web Archive” to new, more focused topically-based web archives.

Animal rights & Veganism (11)

Civil Rights & Human Rights (1)

Economic and Social Justice (3)

Education and Student Movement (8)

Electoral Politics, Political Parties, Political Action (6)

Environmentalism / Green Movement (2) – includes food and agriculture

Feminism & Women’s Movements (6) – includes abortion rights

Internet / Cyberspace Democracy (1)

Notable Individuals (3)

Short term collecting priorities for the above archives:

Enrich all the Library’s not-publicly available web archives to “critical mass,” i.e., to the number of websites for each that will make it appropriate to make these archives public. While this number may vary according to the level or activity and the scope of the topic, it should generally be at least 10-20 websites.

Create the following collections:

Note: *Some of these topics are new, while for some there are already some such websites in the “Other Left Activism” web archive.*

Academic freedom (Frederic Ewen Academic Freedom Center connection)
Globalization
Gay & Lesbian rights
Housing, squatters and tenants movements
Library worker left
Online Social Networks
Philanthropy, Progressive
Progressive “Think Tanks” -- Research / Policy Groups

Develop a File of Websites of Interest, and a List of Prospective Web Archives:

Collection development activity, which by its very nature involves the revelation and exploration of associated URLs, inevitably discovers websites of interest that may be relevant for current web archives, or that may suggest the formation of additional web archives. Maintaining for future reference a file of such URLs, and a list of possible web archives (drawn in part from this file), should be an integral part of the collection development process, since collection development must be “disciplined” by our collecting policy, and not fly off in all directions.

An Archival Approach / the Web Archiving Service:

Since this political website project is being conducted as part of Tamiment’s archival program, the curator will approach selection from an archival perspective. Websites, like archives have an organic unity, and a “look and feel,” which the Library will seek to capture and preserve. For some entities whose web presence we wish to archive, this may require capturing not only the content contained within the root URI, but content from one or more other “parallel” URIs.

An archived website is defined as that which is retrieved by the web crawling software. (For us, this is the Web Archiving Service, open source software developed by the California Digital Library, with “high end” capture and metadata capabilities.) Captured content will include hypertext/media and style sheets that may be activated in order to retrieve additional web-based published materials, including audio and images.

Some types of content are not currently capable of being captured, e.g., pages that rely on embedded databases, and portions of websites whose capture are forbidden by their owners (so-called robot exclusions).

However, Tamiment chose the WAS not only for its “high end” capabilities, but because of its commitment to ongoing improvement to its web crawling, searching, and metadata capabilities, and the WAS has been responsive to the Library’s questions, concerns, and requests for enhancements.

The default capture scope will be the entire website. However, at least annually we will capture the website plus linked pages, as a prominent feature of web sites are links to related content pages, and users will want to be able to view the homepages of the linked sites, to get a basic understanding of the entities that created them and their missions. The default capture frequency is quarterly, however, during important periods in the life of the entities whose websites we capture, say a strike by a labor union, or an election, more frequent (weekly or daily), but briefer captures can be instituted, almost always at the home page + linked pages setting (equivalent to seeing the front page

of a newspaper, and all the articles thereon that extend onto other pages), or sometimes the homepage only.

Archived copies of web sites are made available six months after the date of capture. In essence, this means that websites are viewed as published materials, and the six month delay gives the website owner a chance to opt out of the archive (the WAS software informs the website that it is being crawled, and also honors all websites' robot exclusions).

In-house WAS analytical data is very rich, and includes numerous reports on individual captures, including number and types of files captured and not captured. Reports comparing different captures of the same website show the number of unchanged files, deleted files, modified files, and new files – all obviously important for analyzing an archived website, including the identification of inactive websites.

Publicly accessible website analytical data includes brief descriptive metadata (site name, one sentence description) and technical metadata (original and archival URI, date of capture, size of capture in bytes, and whether the capture was direct or indirect, e.g., the byproduct of a host + linked pages capture of another website).

There are, and will continue to be, finding aids for each web archive, whose “container list” will consist of archived website names and the URL of the earliest capture thereof. Although the Library does not individually catalog websites, within the WAS website description module LC terms are entered in the “subjects” field, in anticipation of a projected upgrade to the WAS' search capabilities to allow key word descriptor searching. Currently, WAS searching is full text (including URLs), either at the individual web archive level, at the level of the set of captures of an individual website (either of these can be further limited by file type, e.g., html, pdf, images, audio, etc.). However, searching cannot yet be limited to a specific capture of a website, or expanded to include all of Tamiment's individual web archives, although the latter functionality should be added in the coming academic year.

Usage, collection and website analysis, researcher feedback, and an ongoing evaluation of the scholarly literature may lead to adjustments in collection guidelines and procedures.

Curator: Peter Meyer Filardo, Tamiment Archivist (for political left collections)

Tamiment Library
New York University
70 Washington Square South
New York, N.Y. 10012
peter.filardo@nyu.edu
212-998-2639

Appendix: Other Left Activism Web Archive

[ABC No Rio](#)
[Abraham Lincoln Brigade: Spanish Civil War History and Education: Abraham Lincoln Brigade Archives](#)
[Acorn](#)
[Act Up NY](#)
[American Constitution Society](#)
[American Friends Service Committee](#)
[American Friends Service Committee Prison Watch](#)
[American Library Association. Social Responsibilities Round Table](#)
[Animal Liberation Front](#)
[Asian American Legal Defense Fund](#)
[Assata Shakur Speaks!](#)
[Behind the Homefront](#)
[Billionaires For Bush](#)
[Black Radical Congress](#)
[Blackout Arts Collective](#)
[Brights' Net, The](#)
[Center for Constitutional Rights](#)
[Center for Constitutional Rights 2](#)
[Center for Popular Economics - CPE](#)
[centerforconstittutionalrights](#)
[CorpWatch](#)
[Critical Resistance](#)
[DRUM Desis Up and Moving](#)
[Econ-Atrocity / Econ-Utopia » Field Guide to the U.S. Economy](#)
[Experimental Party Disinformation Center](#)
[friends of the earth international](#)
[Green Party of the United States](#)
[Greenwich House](#)
[guerrilla girls](#)
[health global access project](#)
[Hope Along the Wind - Harry Hay](#)
[Hudson River Sloop Clearwater](#)
[Immigrant Solidarity Network](#)
[institute for social ecology](#)
[Interactivist Network](#)
[International Endowment for Democracy - supporting democracy in the country that needs it most - the USA](#)
[jubilee usa](#)
[Kucinich: Congressman Dennis Kucinich](#)
[Labor Party \(U.S.\)](#)
[Living Liberally | Political action through social interaction](#)
[madre](#)
[May First/People Link](#)
[Metropolitan Council on Housing \(New York City\)](#)
[Michael Parenti Political Archive](#)
[Nader, Ralph: Page](#)
[Naomi Klein.](#)
[national lawyers guild](#)
[National Organization for Women](#)
[National Organization for Women. New York City.](#)
[National War Tax Resistance Coordinating Committee](#)
[New York Collective of Radical Educators](#)
[New York State Greens/Green Party of New York State](#)
[North Star Fund](#)
[Nyc FTAA Referendum](#)
[People for the American Way.](#)

[proliberated freedom campaign](#)
[Radical Homosexual Agenda](#)
[Radical Philosophy Association](#)
[Radical Reference | Answers for those who question authority](#)
[Reverend Billy & The Church of Life After Shopping](#)
[School of the Americas Watch \(SOAW\)](#)
[Sista II Sista.](#)
[Southern California Library](#)
[Students Against Sweatshops](#)
[Students for a Democratic Society](#)
[Students for a Free Tibet](#)
[Tamiment Library & Robert F. Wagner Labor Archives, New York University.](#)
[The Struggle for Free Speech at CCNY, 1931-42](#)
[TIME'S UP! :: NYC Direct Action Environmental Organization](#)
[Union for Radical Political Economics - URPE](#)
[united for peace and justice](#)
[Vote Rev Billy For Mayor Of NYC](#)
[Welcome to Critical Sociology](#)
[Welcome to MichaelMoore.com](#)
[Welcome to the Official Woody Guthrie Website](#)
[William Kunstler: Disturbing the Universe](#)
[working families party](#)