

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Archiving the Russian and East European Lesbian, Gay, Bisexual, and Transgender Web, 2013: A Pilot Project

### Permalink

<https://escholarship.org/uc/item/23r8r260>

### Journal

Slavic & East European Information Resources, 15(3)

### ISSN

1522-8886 1522-9041

### Author

Pendse, Liladhar R

### Publication Date

2014-08-08

### DOI

10.1080/15228886.2014.930973

Peer reviewed

Archiving the Russian and East European Lesbian, Gay, Bi-sexual and Transgender (LGBT)

Web 2013: A Pilot Project

Liladhar R. Pendse, Ph.D.

University of California, Berkeley

Doe Library

[Lpendse@library.berkeley.edu](mailto:Lpendse@library.berkeley.edu)

**Abstract:**

This paper focuses on conceptualization and implementation of a web archiving pilot project at UC Berkeley Library of selected Russian and East European Lesbian, Gay, Bi-sexual and Transgender websites for academic research purpose. The paper introduces use of the Web Archiving Services (WAS) platform of the California Digital Library of the University of California. While identifying the criteria that were used to harvest these websites, the paper also identifies various complexities that are associated with the viability of projects having complex social and political issues, such as the Russian and Eastern European LGBT rights movements. However, this paper does not take an ideological stance with respect to the legal aspects of such social interactions, but rather strives to preserve information having a purely academic research nature.

**Keywords:** Web Archive, Lesbian, Gay, Bi-sexual and Transgender web, Eastern Europe, Russia.

**Introduction:**

The purpose of this pilot project is to explore meaningful strategies and ways that can be used to preserve some portion of the Russian and Eastern European LGBT web in light of constantly evolving Russian and East European legal and social attitudes towards various alternate lifestyles. This pilot project began in July of 2013 and is planned to end in January of 2014.

It is a common knowledge that many of the born digital objects and their reiterations on the World Wide Web (the web) become unusable and extinct due to multiple reasons, like rapidly changing technologies, inability of their creators to maintain these sites over a long

period of time on the Web, and other reasons such as legal impediments to maintaining such websites in the different countries.<sup>1</sup> I have used the term web-archiving to summarize our activities for culling, preserving and archiving the selected LGBT web-sites for the future under the premise that the digital information that these contain has both academic and cultural values. The LGBT archive per say is not a new term. For example, One National Gay & Lesbian Archives, currently housed in the University of Southern California Library, was founded in 1952.<sup>2</sup> There are several other international examples of archiving print and non-print materials, such as the Canadian Lesbian+Gay Archives or *Internaational home/lesbisch Informatiecentrum en archief* (IHLIA) in the Netherlands.<sup>3</sup> Although the above mentioned archives collect materials related to the LGBT movement, they do not have systematic web-archiving policies

---

<sup>1</sup> Denber, Rachel. "President Putin's First 60 Days; the Kremlin Lays the Legal Groundwork for Deeper Authoritarianism in Russia." *Wall Street Journal (Online)* Jul 16 2012. *ProQuest*. Web. 20 Oct. 2013. Rachel Denber quotes, "*Meanwhile, Mr. Medvedev once promoted the Internet as crucial to Russia's economic development. But under a draft law recently introduced, Russian authorities would have the power to draw up a list of Web content that they deem harmful to children or that any court has found to be "extremist."*

*The host of any website found to contain such content would have 24 hours to notify the site's owner, who in turn would have 24 hours to remove the content. If the owner failed to do so, the Web host would have to block the entire site. This continues efforts that have been under way at the regional level for some time.*

*Russian authorities are known to use other laws restricting expression, particularly those on "extremism," to crack down on dissent and could use the draft law to shut down protest-group websites. Also, now that several Russian provinces and cities have adopted laws banning "homosexual propaganda"--allegedly to protect children--and a relevant federal draft law has also been proposed, Web content on gay and lesbian issues could be seriously jeopardized.*

<sup>2</sup> "ONE National Gay & Lesbian Archives. The Independent, Community-Partner of ONE Archives at the USC Libraries." *ONE National Gay Lesbian Archives*. One National Gay & Lesbian Archives, 2012. Web. 20 Oct. 2013.

<sup>3</sup> "Zoeken & Vinden." *IHLIA*. N.p. Web. 22 Oct 2013. <<http://www.ihlia.nl/>>.

that would preserve the LGBT web for posterity. In addition, to our knowledge, there hasn't been any prior systematic pilot project of archiving Russian and East European LGBT web in an academic setting in the United States.

Alternate lifestyles have been a contentious issue in contemporary Eastern and Russian society.<sup>4</sup> On one hand, there is a level of tacit acceptance of alternate life styles by some, However, there are several laws in place that outlaw homosexuality.<sup>5</sup> These lifestyles have tended to spark societal debates ranging from polemical wrangling to violent protests against the LGBT community.<sup>6</sup> Without going in the details of the debates surrounding the recently passed legislation that makes "propaganda of nontraditional sexual relations among minors" illegal, one would argue that the overall political climate in some of the Eastern European countries and Russia with respect to LGBT expression will remain problematic. This, in turn, might negatively impact the viability and sustainability of information that is contained on the web-sites of LGBT organizations. Thus, the desire to preserve and protect this potentially endangered information on LGBT community affairs, information created by the community itself for future academic and research needs, was the main purpose behind this pilot project. Besides, this primary goal, we also wanted to test the capabilities of the different web-archiving software platforms and see how each platform can be adapted to our needs for web-archiving.

### **Literature Review:**

---

<sup>4</sup> *Russian Gays/western Gaze: Mapping (homo)sexual Desire in Post-Soviet Russia*. Duke University Press, 2002. Internet resource.

<sup>5</sup> Waaldijk, C. *Tip of an Iceberg: Anti-lesbian and Anti-Gay Discrimination in Europe, 1980-1990 : a Survey of Discrimination and Anti-Discrimination in Law and Society*. Utrecht: International Lesbian and Gay Association, 1991. Print.

<sup>6</sup> "Gay Russians are 'Free' Now but Still Stay in Fearful Closet." *New York Times (1923-Current file)*: 1. Jul 08 1995. *ProQuest*. Web. 20 Oct. 2013 .

Much attention has been paid to social, health and political issues surrounding LGBT communities in the Eastern Europe and Russia<sup>7</sup>.<sup>8</sup> However, academic level works that analyze the LGBT web in the region are very limited and virtually non-existent.<sup>9</sup> On the other hand, there is a whole plethora of literature on the preservation of the web-sites. For example, just a simple keyword search in Worldcat for the term “web-archiving” produces several hundred results. Much attention has been paid to the social network analysis and how the websites link from and to each other.<sup>10</sup> There have been several efforts geared towards the preservation of the LGBT Web.<sup>11</sup> Web-archiving has been successfully implemented using various software platforms by multiple institutions. There are multiple commercial and non-commercial open-source software packages that can be used to harvest a particular website.<sup>12</sup> Besides individual organizational efforts, one has access to existing projects like the Wayback Machine or Internet Archive.<sup>13</sup> Other academic works argue for web-archiving and cite multiple reasons that highlight the

---

<sup>7</sup> Takács, Judit, and Ivett Szalma. "Homophobia and Same-Sex Partnership Legislation in Europe." *Equality, Diversity and Inclusion: an International Journal*. 30.5 (2011): 356-378. Print.

<sup>8</sup> Niccolai, LM, EJ King, KU Eritsyan, L Safiullina, and MM Rusakova. "'in Different Situations, in Different Ways': Male Sex Work in St. Petersburg, Russia." *Culture, Health & Sexuality*. 15.4 (2013): 480-93. Print.

<sup>9</sup> Barriault, Marcel. "Hard to Dismiss: The Archival Value of Gay Male Erotica and Pornography." *Archivaria*.68 (2009): 219-46.*ProQuest*. Web. 21 Oct. 2013.

<sup>10</sup> Bredl, Klaus, Julia Hünninger, and Jakob L. Jensen. *Methods for Analyzing Social Media*. London: Routledge, 2014. Print. Faheem, M, and P Senellart. "Intelligent and Adaptive Crawling of Web Applications for Web Archiving." *Lecture Notes in Computer Science*. (2013): 306-322. Print.

<sup>11</sup> Lynne Kirste. "Collective Effort: Archiving LGBT Moving Images." *Cinema Journal* 46.3 (2007): 134-140.*Project MUSE*. Web. 27 Oct. 2013. <<http://muse.jhu.edu/>>.

<sup>12</sup> "Tools and Software." *International Internet Preservation Consortium (IIPC)*. N.p., n.d. Web. 21 Oct. 2013.

<sup>13</sup> "About the Wayback Machine." *Internet Archive: Wayback Machine*. N.p., n.d. Web. 21 Oct. 2013.

necessity for web-archiving. These reasons vary from the definition of the web as a distributed information repository to organizational recordkeeping mandates.<sup>14</sup>

**Background:**

The idea of the Russian and Eastern European Web-archiving project originated during my work as the librarian for Slavic and East European collections at the Princeton University's Firestone Library. I had subscribed for our collections issues of a Russian gay magazine called Kvir.

See:

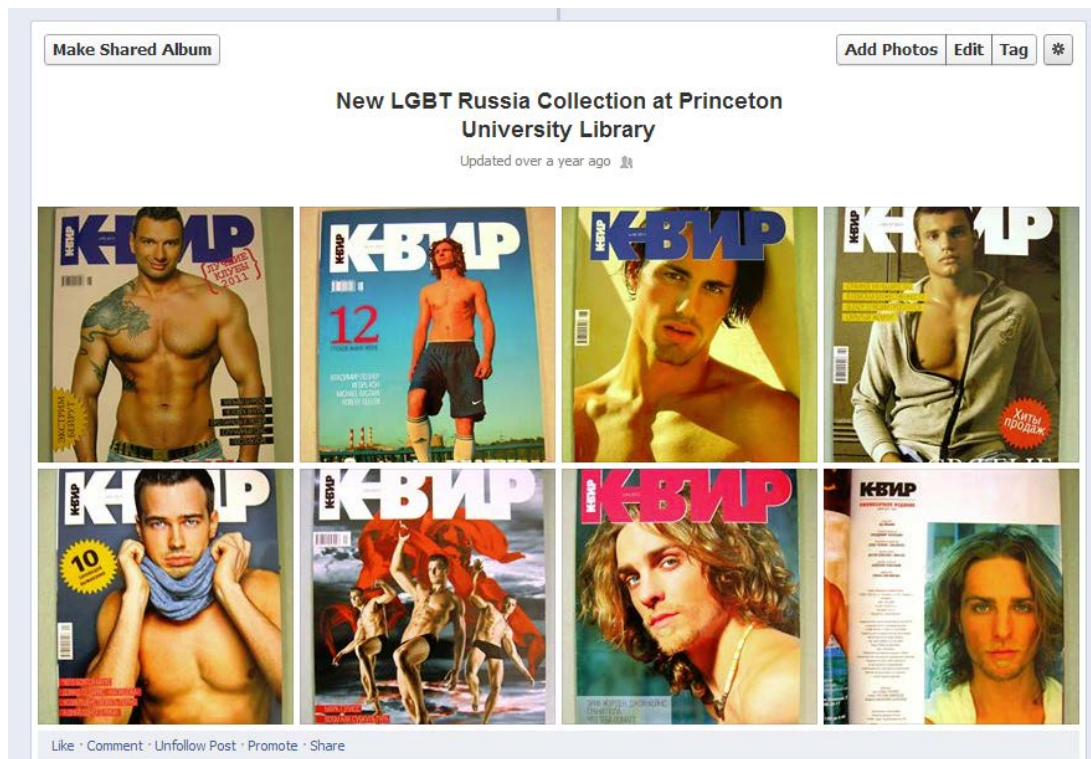
<https://www.facebook.com/media/set/?set=a.208031782640812.42921.100003019757366&type=1&l=9974da1fc6>

---

<sup>14</sup> Oury, Clement, Tobias Steinke, and Gina Jones. "Ensuring Long-Term Access to the Memory of the Web Preservation Working Group of the International Internet Preservation Consortium." *International Preservation News*.58 (2012): 34-7. *ProQuest*. Web. 21 Oct. 2013.

Farmer, Lesley S. J. "The Life Cycle of Digital Reference Sources." *Reference Librarian* 50.2 (2009): 117-36. *ProQuest*. Web. 21 Oct. 2013.

Figure 1: Kvir Journal Collection at Princeton University Library.



However, I could not get all of the back issues. The vendor assured me that these were available online, yet I could not find these issues displayed online. In my conversations with other faculty members, and colleagues at UCLA, Princeton and UC Berkeley, I realized one way to provide access to the ever-changing landscape of the LGBT websites from the region would be to implement a pilot project for archiving the LGBT web. Thus, I started to identify a representative group of the LGBT websites from the region. Besides this, I also used both HTTrack and Wecopier platforms harvesting the data.<sup>15</sup> Both of these software platforms at that time had limitations. The data that were harvested could be stored only on the local hard-drive. Also, the automatic harvesting features on these both software platforms were very limited.

<sup>15</sup> "HTTrack Website Copier." *HTTrack*. N.p., n.d. Web. 27 Oct 2013. <<http://www.httrack.com/>>.

Some of my colleagues suggested using the Wayback machine to obtain the data. The following picture perhaps can shed some light on how the Wayback machine does not necessarily capture or provide a reliable access to the LGBT websites.

Figure 2: Gayrussia.ru webpage captures on the Wayback Machine.

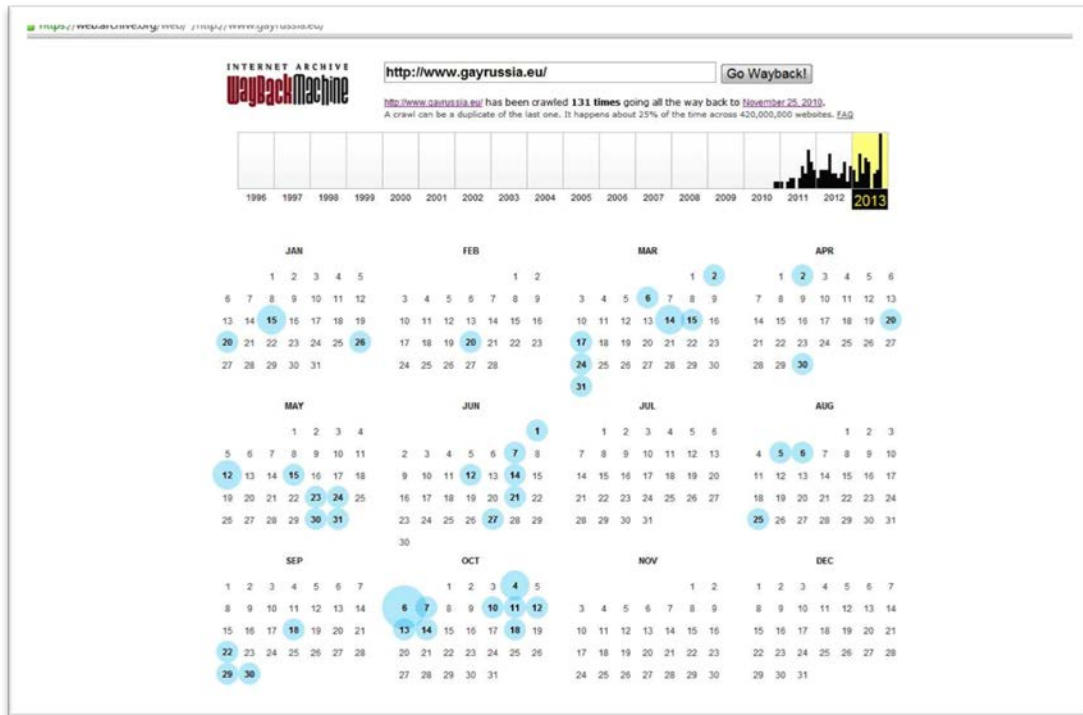


Figure 3: The capture for November 25, 2010 does not indicate any content.

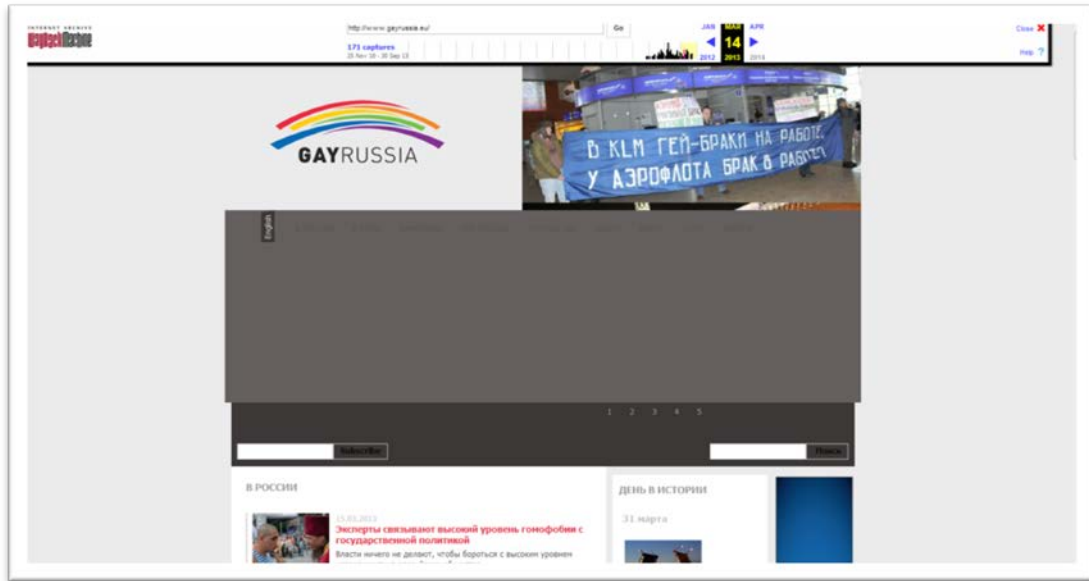


One particular instance of lack of access to the content of this website should not be generalized to say that the Wayback Machine does not properly archive these types of web-sites.



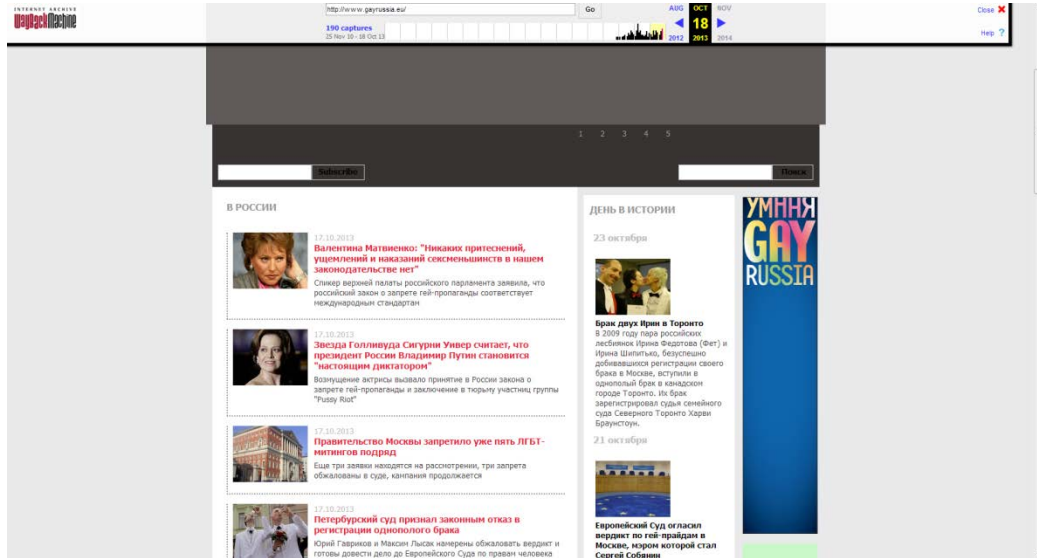
Indded, later archival captures in 2012 and 2013 do provide access to the site, and also to the content changes as these evolve. Nevertheless, the Wayback machine does not allow us to automatically compare files that change or even disappear. Also the Wayback machine only collects the data that is considered to be from the surface of the web.<sup>16</sup>

Figure 4: A screen-shot of a capture for 14 March 2013.



<sup>16</sup> Lecher, Hanno E. (2006). Small scale academic web archiving: DACHS. In *Web Archiving* (pp. 213-226). Berlin: Springer-Verlag.

Figure 5: A capture for 18 October 2013.



While recognizing the importance of the Wayback machine as a general tool for Web preservation, I decided to see if there were any other alternate ways in which I could preserve selected portions of the Eastern and Central European LGBT web. In consultation with other colleagues at UC Berkeley, I decided to use a centralized platform of the University of California called the Web Archiving Service (WAS).

### **The WAS model of the University of California:**

The University of California's Web Archiving Service (WAS) was established as a web archiving service in order to enable various institutes to create enduring digital archives of fragile web resources and safeguard them in long-term private or publically accessible storage. WAS is a service of the University of California Curation Center, powered by the California Digital Library (CDL).<sup>17</sup> WAS is uniquely positioned to aid libraries and institutions in the web

<sup>17</sup> "California Digital Library." *California Digital Library*. The Regents of The University of California, n.d. Web. 27 Oct. 2013.

archiving process. We understand the varieties of research and types of materials you may want to collect, and the importance of integration with your existing systems.

The WAS allow its users to collect and preserve hundreds of hard-to-find and continually updated web pages, related sites, and sub-sites on a periodic schedule that can be automated. One feature of the WAS, that I appreciated the most, was to have an ability to create new collections. The WAS allows users or curators to collect and preserve web-sites that are linked by common themes or subjects.

The WAS interface is user-friendly as it allows the user or the creator of the content to enter the URLs one wants, and the WAS web crawlers then can harvest the data. There are both enhanced customer support, and as-needed guides, and in-depth training sessions.

WAS hosts the service and the storage. CDL hosts both the service and storage, freeing the user of the need for additional storage infrastructure or dedicated IT staff.

Figure 6: California Curation Center.

The screenshot shows the website for the University of California Curation Center. At the top, there is a navigation bar with links for Staff Directory, Contact CDL, Report a Problem, and System Status. Below this is a search bar and a secondary navigation menu with categories like About CDL, Services and Projects, Information Gateways, Committees and Groups, and News and Media. The main content area is titled "University of California Curation Center" and includes a brief description of UC3's role. It features three main service sections: Merritt (a repository service), EZID (a service for creating and managing identifiers), and Web Archiving Service (for capturing and archiving web content). On the right side, there is a "News" sidebar with a list of recent updates, including Data Citation Developments, RDA Meeting updates, and various service updates for Merritt, EZID, and WAS.

## Methodology:

For this pilot project I was able to initially identify fifteen LGBT websites from the Eastern Europe and Russia. Towards this end, I used purposive sampling and conducted search for these websites using popular search engines. The sites were chosen individually for their content. We excluded e sites with pornographic content or ones that were geared towards dating. The purpose of this project was to archive the websites that contained information relevant to LGBT movements in Russia and the Eastern Europe. I also used the directory of the International Lesbian, Gay, Bisexual and Allied Organizations' (ILGA) directory.<sup>18</sup>

Figure 7: An example of Lesbi.ru was chosen for harvesting of the data.

<sup>18</sup> "International Lesbian, Gay, Bisexual, Trans and Intersex Association.

<http://ilga.org/directory/en/ILGA>, n.d. Web. 24 Oct. 2013.

The criteria for choosing these sites were developed as follows,

1. Name of the Site and the name of Organization that is the sponsor for this site.
2. Is this a governmental or NGO site?
3. How is the content of the site that is chosen relevant to research needs of the futures scholars?
4. What type of interactive features such as music, clips or archival files are embedded in the selected website?
5. Do these sources contain within specific historical, cultural, social, educational, or academic viewpoints?<sup>19</sup>

The selection criteria were further modified to suit the needs of this particular pilot project along the lines that the National Archives and Records Administration has recommended. The U.S. National Archives and Records Administration (NARA) had deployed several units of analysis in its guidance for government agencies to conduct risk analysis for web records: group of websites, a whole website, a website minus one or two portions that exhibit substantially different characteristics, and clusters of web pages (NARA, 2005).<sup>20</sup> However, we recognize that the LGBT websites of the region are not necessarily governmental websites.

The sites that I had identified for this project were as follows,

---

<sup>19</sup> Chen K. H., Chen, Y. L., & Ting, P. F. (2008). [Developing National Taiwan University Web Archiving System](#). Proceedings from IWWA '08: *8th International Workshop for Web Archiving*, Denmark (pp. 1-8).

<sup>20</sup> NARA. (2005). [NARA Guidance on Managing Web Records](#). NARA.

Table 1: The list of the LGBT sites that were chosen for this pilot project.

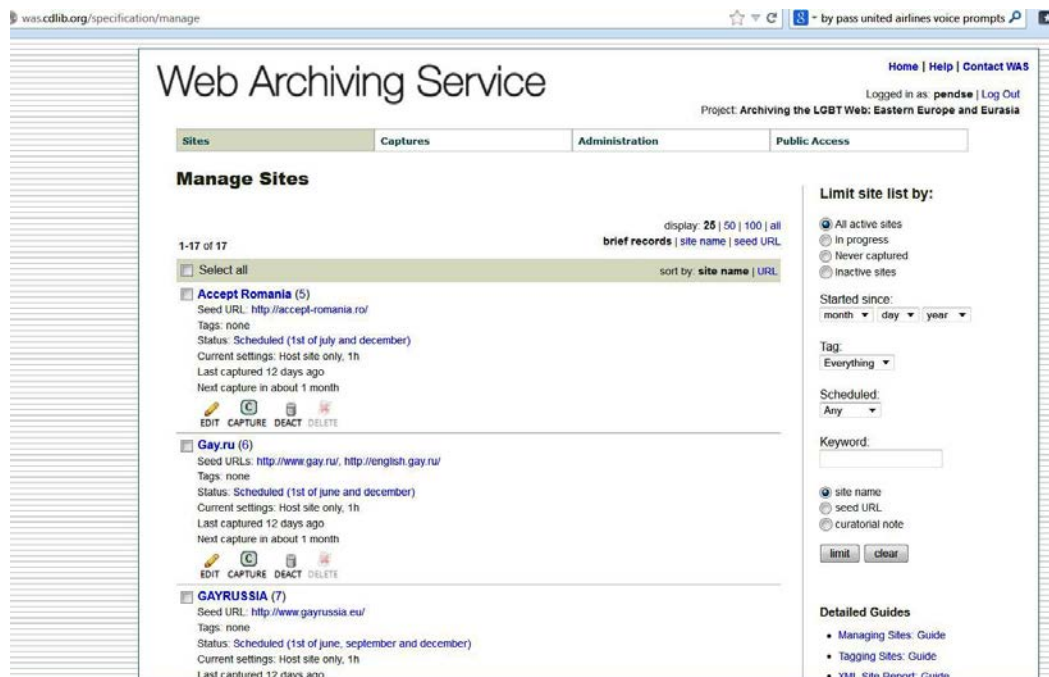
Title	Country	Web link
Accept Romania	Romania	<a href="http://accept-romania.ro/">http://accept-romania.ro/</a>
Gay Portal Ukrainy	Ukraine	<a href="http://gayua.com/main.html#">http://gayua.com/main.html#</a>
Lesbiru.com	Russian Federation	<a href="http://lesbiru.com/">http://lesbiru.com/</a>
Sunshine Belarus	Belarus	<a href="http://sunshine.by/">http://sunshine.by/</a>
Szerbiai Magyar LMBT Csoport	Serbia	<a href="http://szmlmbt.blogspot.com/">http://szmlmbt.blogspot.com/</a>
Lygių Galimybių Pletros Centras	Lithuania	<a href="http://www.gap.lt/main.php/id/1">http://www.gap.lt/main.php/id/1</a>
Istorak	Croatia	<a href="http://www.gay.hr/">http://www.gay.hr/</a>
Gay.Ru	Russian Federation	<a href="http://www.gay.ru/">http://www.gay.ru/</a>
Gay Russia	Russian Federation	<a href="http://www.gayrussia.eu/">http://www.gayrussia.eu/</a>
Inakost'	Slovakia	<a href="http://www.inakost.sk/">http://www.inakost.sk/</a>
Lezbijska Grupa Kontra	Croatia	<a href="http://www.kontra.hr/cms/index.php?option=com_content&amp;view=article&amp;id=15&amp;Itemid=45&amp;lang=en">http://www.kontra.hr/cms/index.php?option=com_content&amp;view=article&amp;id=15&amp;Itemid=45&amp;lang=en</a>
Kvir.ru	Russian Federation	<a href="http://www.kvir.ru/">http://www.kvir.ru/</a>
Lesba.sk	Slovakia	<a href="http://www.lesba.sk/">http://www.lesba.sk/</a>
Lesbi.ru	Russian Federation	<a href="http://www.lesbi.ru/">http://www.lesbi.ru/</a>

Web Archiving Service’s interface is simple and intuitive. See the screenshot below:

Figure 8: The WAS Interface

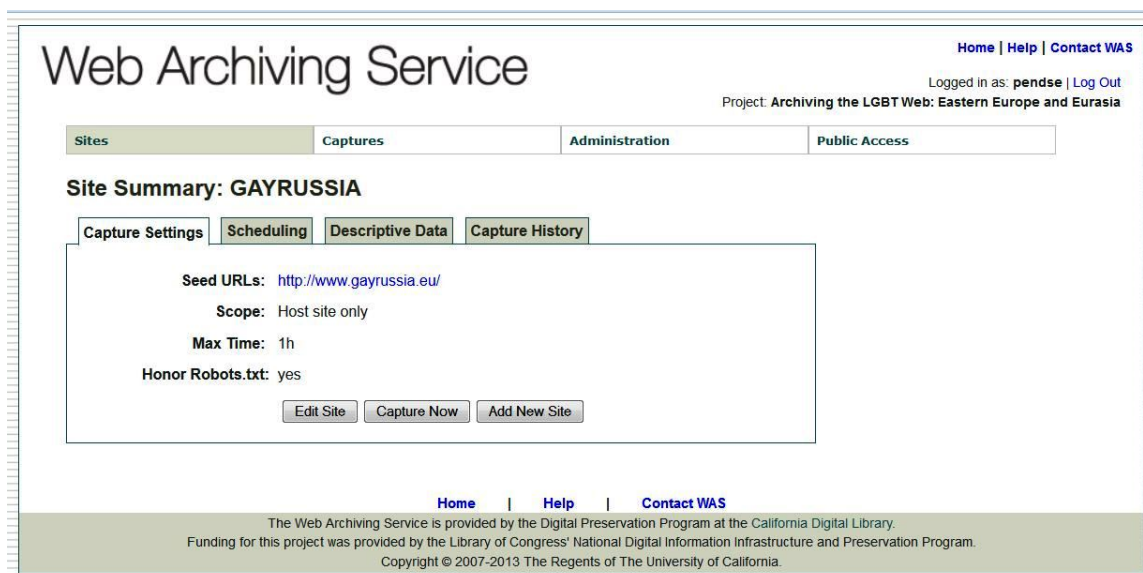


Figure 9 The site management component.



The interface as I had mentioned is very intuitive and for each archived site the archiving interval can be changed as shown in the image below. The site's capture can be deactivated and reactivated if the need arises. These captures can also be permanently deleted.

Figure 10: Capture Settings, Scheduling, Descriptive data and Capture History.



The capture settings allow the curator of particular thematic collection to choose the duration for a particular capture. There are two duration settings, i.e., for one hour and for twenty-four hours. The capture setting prompts the curator to honor the robots.txt or not to honor it.

Web site owners use the /robots.txt file to give instructions about their site to web crawling robots such as those of the WAS; this is called The Robots Exclusion Protocol. It works like this: a robot wants to visit a Web site URL, say <http://www.example.com/welcome.html>. Before it does so, it firsts checks for <http://www.example.com/robots.txt>, and finds:

User-agent: \*

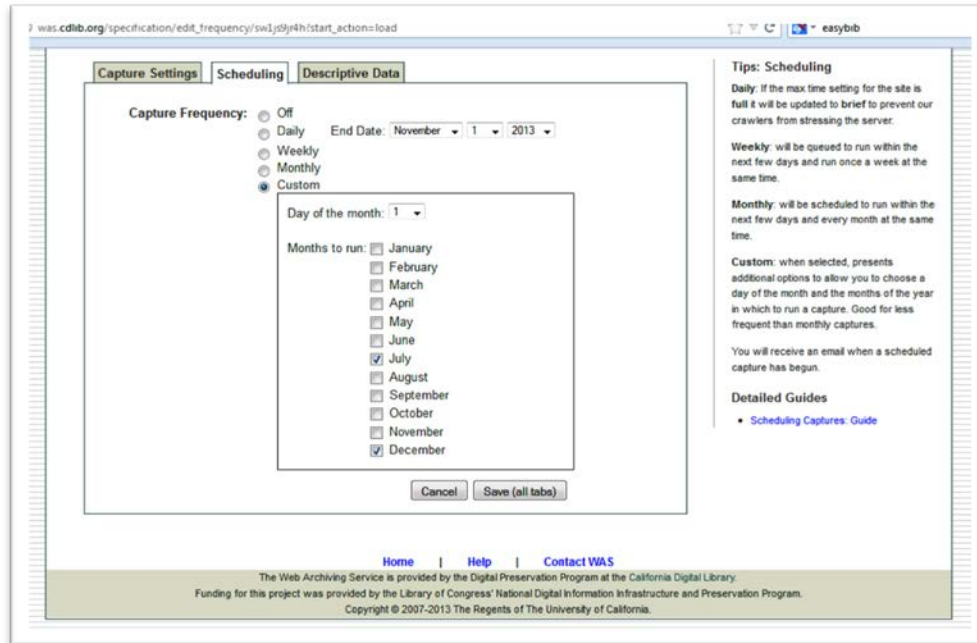
Disallow: /



The "User-agent: \*" means this section applies to all robots. The "Disallow: /" tells the robot that it should not visit any pages on the site.<sup>21</sup> This way the robots are unable to crawl and preserve the website. However the WAS allows the user to override the robots.txt and allow the curator to cull the data.

The other important feature of the WAS is automated scheduling. One can capture the chosen websites on daily, weekly, and monthly basis. There is also a custom option that allows the curator to choose specific months and frequency of captures.

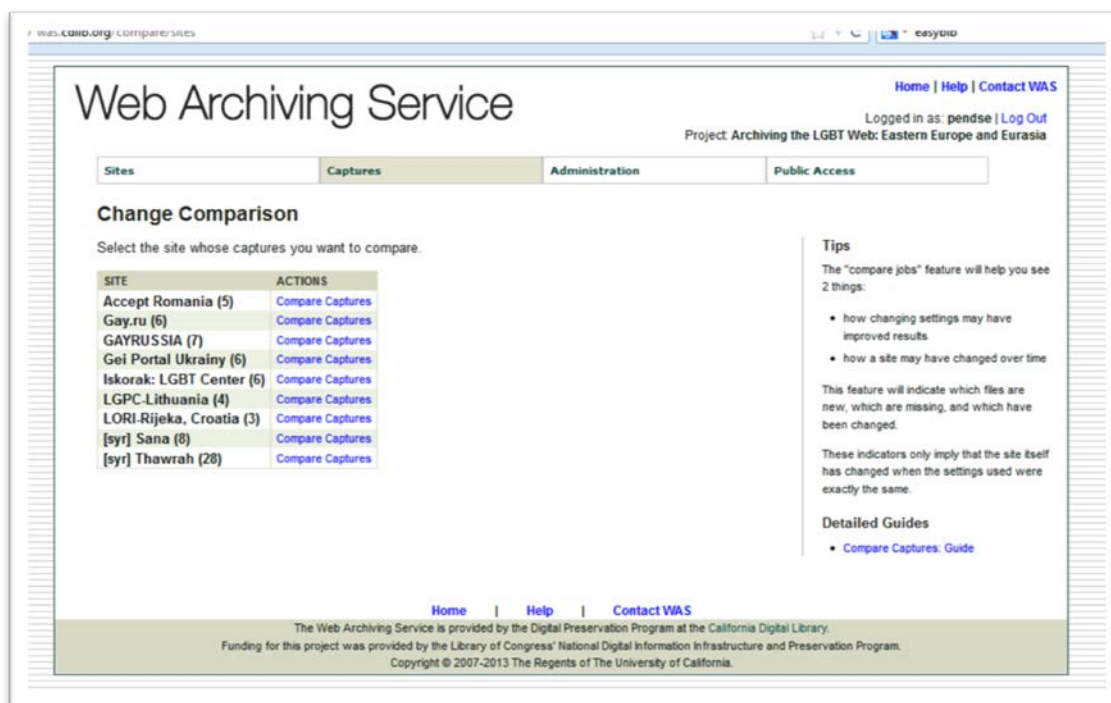
Figure 11: Custom capture setting menu.



The WAS allows its user to compare various captures. This in turn helps us to identify the files that were either added or have been replaced since the last crawl.

<sup>21</sup> "Robotstext.org." N.p., 2007. Web. 25 Oct. 2013. <<http://www.robotstxt.org/robotstxt.html>>.

Figure 12: Change comparison of captures.



The quantity of files captured indicates one of two things. First, how changing settings may have improved results. Second, how a site may have changed over time. This feature also indicates which files are new, which are missing, and which have been changed. Let us quickly see an example of Iskorak.

Iskorak or Breakthrough - the Center for Sexual and Gender minorities is an NGO that was founded in Croatia on 12 January 2002. It is committed to fight against all forms of discrimination and stigmatization of sexual minorities. This NGO fights for the increased participation in society as equal citizens of the Croatia.

Figure 13: Iskorak.hr

www.iskorak.hr

Dva od tri gej, bi i maskarada iz Zagreba **nisu se testirali u zadnjih godinu dana.**

**A svaki drugi HIV+ ne zna da je pozitivan.**

**Znaš li ti svoj HIV status?**

NAŠI PROJEKTI

- Aids hr
- Split pride
- Pravni tim

Figure 14: Processing of the changed files for Iskorak center.

was.cdlib.org/compare/wait\_message?jobs%5B%5D=00364534&jobs%5B%5D=00413665&referer=%2Fcompare%2Fcompare\_jobs%3Fjob\_00413665%

Web Archiving Service

Home | Help | Contact WAS

Logged in as: pendse | Log Out

Project: Archiving the LGBT Web: Eastern Europe and Eurasia

Sites Captures Administration Public Access

**Processing comparison information for Iskorak: LGBT Center**

The captures you selected are being processed for comparison. Once the progress bar reaches 100% you will be taken to the comparison screen for these captures.

Processed 1,102 of 4,529 files

Home | Help | Contact WAS

The Web Archiving Service is provided by the Digital Preservation Program at the California Digital Library. Funding for this project was provided by the Library of Congress' National Digital Information Infrastructure and Preservation Program. Copyright © 2007-2013 The Regents of The University of California.

Iskorak's two captures and various changed files are indicated below in the figure 14.

Figure 15: Iskorak's changed files.

The screenshot shows the WAS platform interface for comparing file captures. The main content area displays a 'Change Summary' table with the following data:

Change Summary
Changed: 238 files (2.1% by size)
New: 1,571 files (41.9% by size)
Missing: 2,082 files (49.3% by size)
Unchanged: 200 files (6.7% by size)

Below the summary is a list of files with their URLs and sizes, sorted by URL. The list includes:

URL	File Size
<a href="http://24.media.tumblr.com/80da624510a95...654a/tumblr_mi7uayapnO1r43g5ko1_500.jpg">http://24.media.tumblr.com/80da624510a95...654a/tumblr_mi7uayapnO1r43g5ko1_500.jpg</a>	67.2 kB
<a href="http://25.media.tumblr.com/d874208653c2a...d0f7/tumblr_mok7799VAY1q9ytzo1_1280.jpg">http://25.media.tumblr.com/d874208653c2a...d0f7/tumblr_mok7799VAY1q9ytzo1_1280.jpg</a>	139.5 kB
<a href="http://i.imgur.com/1UU5148.jpg">http://i.imgur.com/1UU5148.jpg</a>	30.3 kB
<a href="http://i.imgur.com/3DruzEV.jpg">http://i.imgur.com/3DruzEV.jpg</a>	59.8 kB
<a href="http://i.imgur.com/9SS9IKg.jpg">http://i.imgur.com/9SS9IKg.jpg</a>	4.8 kB
<a href="http://i.imgur.com/dgWwPhz.jpg">http://i.imgur.com/dgWwPhz.jpg</a>	62.5 kB
<a href="http://i.imgur.com/g2xHvjq.jpg">http://i.imgur.com/g2xHvjq.jpg</a>	20.9 kB
<a href="http://i.imgur.com/huPhz6.jpg">http://i.imgur.com/huPhz6.jpg</a>	51.7 kB
<a href="http://i.imgur.com/iAWIT7K.jpg">http://i.imgur.com/iAWIT7K.jpg</a>	32.9 kB
<a href="http://i.imgur.com/kzTr8Zk.jpg">http://i.imgur.com/kzTr8Zk.jpg</a>	32.6 kB
<a href="http://i.imgur.com/lhniEv.jpg">http://i.imgur.com/lhniEv.jpg</a>	13 kB
<a href="http://i.imgur.com/MVO1Ft1.jpg">http://i.imgur.com/MVO1Ft1.jpg</a>	56.6 kB
<a href="http://i.imgur.com/qmi5GZn.jpg">http://i.imgur.com/qmi5GZn.jpg</a>	45.2 kB
<a href="http://i.imgur.com/QoajSuw.jpg">http://i.imgur.com/QoajSuw.jpg</a>	65 kB
<a href="http://i.imgur.com/robots.txt">http://i.imgur.com/robots.txt</a>	1 Byte
<a href="http://i.ytimg.com/vi/3ROXTFkcf0/hqdefault.jpg">http://i.ytimg.com/vi/3ROXTFkcf0/hqdefault.jpg</a>	15.7 kB
<a href="http://i.ytimg.com/vi/5fm8TAlew0/hqdefault.jpg">http://i.ytimg.com/vi/5fm8TAlew0/hqdefault.jpg</a>	11.7 kB
<a href="http://i.ytimg.com/vi/6fmgYNGNeE/hqdefault.jpg">http://i.ytimg.com/vi/6fmgYNGNeE/hqdefault.jpg</a>	15.7 kB

The sidebar on the right contains 'Tips' and 'Detailed Guides' sections. The 'Filetype Limit' section shows a list of file types: HTML, PDF, Image, MS-Office, Video, Audio, and Compressed. A 'Limit' button is visible below the list.

The WAS platform offers summary of what files have changed while comparing two captures against each other. For example, in the figure 14, we see that a total of 238 of files have changed. There were 42% new files<sup>22</sup> and 49% of files that had in a previous capture have been removed. The Iskorak's site had only 7% of files that had remained unchanged.

While the number of files that change with time is normal for any web-site, the WAS allows us to exactly compare these two or more captures by pointing out precisely the nature of this change.

<sup>22</sup> This figure was rounded up to the nearest two decimal points. Thus 41.9% has been counted as 42% and 49.3% is counted as 49% respectively.

The last feature of the WAS platform that I would like to highlight as of utility for my project is that of the ability of the curator to create succinct descriptive data and make this archive publically available.

The descriptive data elements are self-explanatory and depicted below:

Figure 16: WAS Descriptive Data Interface.<sup>23</sup>

The screenshot shows the 'Web Archiving Service' interface for editing descriptive data. The browser address bar shows 'was.cdlib.org/specification/edit\_descriptive'. The page title is 'Web Archiving Service' and the project is 'Archiving the LGBT Web: Eastern Europe and Eurasia'. The user is logged in as 'pendse'. The interface has a navigation menu with 'Sites', 'Captures', 'Administration', and 'Public Access'. The 'Create Site' section is active, with sub-tabs for 'Capture Settings', 'Scheduling', and 'Descriptive Data'. The 'Descriptive Data' tab contains the following fields:

- Site Description: [Text area]
- Corporate creator: [Text input]
- Personal creator: [Text input]
- Publisher: [Text input]
- Subjects: [Text area]
- Geographic coverage: [Text input]
- Curatorial Notes: [Text area]

Below the Curatorial Notes field, a red warning message states: 'Curatorial notes will not display to the public.' At the bottom of the form are 'Cancel' and 'Save (all tabs)' buttons. On the right side, there is a 'Tips: Descriptive Data' section explaining that tags are for site lists and subjects are for descriptive data. Below that is a 'Detailed Guides' section with links to 'Descriptive Data Guide' and 'Tagging Sites'.

### Future Plans:

The purpose of this paper, as indicated right from the onset, was to see what software platforms are available to us as the end users that will enable us to save a portion of the Russian and Eastern European Lesbian, Gay, Transgender and Bisexual web for future academic research

<sup>23</sup> The WAS doesn't specify which of the many Metadata standards was chosen by the California Digital Library. However, the California Digital Library's website provides the in depth information about the DMP Tool and also outlines its metadata standards.

purposes. This particular pilot project has been designed to have duration of 6 months. There were several challenges in implementing this project that varied from defining the selection criteria for the candidate websites to archiving of the streaming media. To our knowledge, there currently isn't any integrated platform that seamlessly downloads and preserves all of the files of a particular website. Despite these challenges, this pilot project remains one of its kind, geared towards preserving cultural, social and political memories of LGBT groups of Russia and Eastern Europe for the posterity.

### Works Cited

- "About the Wayback Machine." *Internet Archive: Wayback Machine*. N.p., n.d. Web. 21 Oct. 2013.
- "California Digital Library." *California Digital Library*. The Regents of The University of California, n.d. Web. 27 Oct. 2013.
- "Gay Russians are 'Free' Now but Still Stay in Fearful Closet." *New York Times (1923-Current file)*: 1. Jul 08 1995. *ProQuest*. Web. 20 Oct. 2013 .
- "HTTrack Website Copier." *HTTrack*. N.p., n.d. Web. 27 Oct 2013. <<http://www.httrack.com/>>.
- "International Lesbian, Gay, Bisexual, Trans and Intersex Association.  
<http://ilga.org/directory/en/>ILGA, n.d. Web. 24 Oct. 2013.
- "ONE National Gay & Lesbian Archives. The Independent, Community-Partner of ONE Archives at the USC Libraries." *ONE National Gay Lesbian Archives*. One National Gay & Lesbian Archives, 2012. Web. 20 Oct. 2013.
- "Robotstxt.org." N.p., 2007. Web. 25 Oct. 2013. <<http://www.robotstxt.org/robotstxt.html>>.
- "Tools and Software." *International Internet Preservation Consortium (IIPC)*. N.p., n.d. Web. 21 Oct. 2013.

"Zoeken & Vinden." *IHLIA*. N.p. Web. 22 Oct 2013. <<http://www.ihlia.nl/>>.

Barriault, Marcel. "Hard to Dismiss: The Archival Value of Gay Male Erotica and Pornography." *Archivaria*.68 (2009): 219-46.*ProQuest*. Web. 21 Oct. 2013.

Bredl, Klaus, Julia Hünninger, and Jakob L. Jensen. *Methods for Analyzing Social Media*. London: Routledge, 2014. Print. Faheem, M, and P Senellart. "Intelligent and Adaptive Crawling of Web Applications for Web Archiving." *Lecture Notes in Computer Science*. (2013): 306-322. Print.

Chen K. H., Chen, Y. L., & Ting, P. F. (2008). [Developing National Taiwan University Web Archiving System](#). Proceedings from IWWA '08: *8th International Workshop for Web Archiving*, Denmark.

Denber, Rachel. "President Putin's First 60 Days; the Kremlin Lays the Legal Groundwork for Deeper Authoritarianism in Russia." *Wall Street Journal (Online)* Jul 16 2012. *ProQuest*. Web. 20 Oct. 2013.

Farmer, Lesley S. J. "The Life Cycle of Digital Reference Sources." *Reference Librarian* 50.2 (2009): 117-36. *ProQuest*. Web. 21 Oct. 2013.

Lecher, Hanno E. (2006). Small scale academic web archiving: DACHS. In *Web Archiving* (pp. 213-226). Berlin: Springer-Verlag.

Lynne Kirste. "Collective Effort: Archiving LGBT Moving Images." *Cinema Journal* 46.3 (2007): 134-140.*Project MUSE*. Web. 27 Oct. 2013. <<http://muse.jhu.edu/>>.

NARA. (2005). [NARA Guidance on Managing Web Records](#). NARA.

Niccolai, LM, EJ King, KU Eritsyana, L Safiullina, and MM Rusakova. "'in Different Situations, in Different Ways': Male Sex Work in St. Petersburg, Russia." *Culture, Health & Sexuality*. 15.4 (2013): 480-93. Print.

Oury, Clement, Tobias Steinke, and Gina Jones. "Ensuring Long-Term Access to the Memory of the Web Preservation Working Group of the International Internet Preservation Consortium." *International Preservation News*.58 (2012): 34-7. *ProQuest*. Web. 21 Oct. 2013.

*Russian Gays/western Gaze: Mapping (homo)sexual Desire in Post-Soviet Russia*. Duke University Press, 2002. Internet resource.

Takács, Judit, and Ivett Szalma. "Homophobia and Same-Sex Partnership Legislation in Europe." *Equality, Diversity and Inclusion: an International Journal*. 30.5 (2011): 356-378. Print.

Waldijk, C. *Tip of an Iceberg: Anti-lesbian and Anti-Gay Discrimination in Europe, 1980-1990 : a Survey of Discrimination and Anti-Discrimination in Law and Society*. Utrecht: International Lesbian and Gay Association, 1991. Print.