

Proposal

Archiving Websites Related to African Countries in Conflict: a Pilot Project

Submitted by: Marion Frank-Wilson, Karen Fung, Tim Johnson, Lauris Olson, Jason Schultz, Mohamed el Seoud.

Background/context:

- A vast amount of knowledge and information of interest for research and scholarly purposes is produced and disseminated via the world wide web (on websites, in blog posts, on social media sites such as Facebook and Twitter, etc.). This is certainly true in the African context: discussions, analyses, primary information ranging from new literary initiatives to government information and latest news on unfolding political events are all published on the web. Librarians have for some time recognized the need to preserve this sometimes ephemeral information for future generations of researchers but, so far, only sporadic attempts to archive African and/or Africa related websites exist. This proposal will serve as a pilot project which will focus on one subject, i.e., African countries in conflict. With this pilot we want to explore the feasibility to create a more extensive portal of African/Africa-related websites at a later point (assuming funding can be found).
- In addition to preserving web resources of interest for this pilot project, the Archive-It subscription service will allow us to build a collection of websites and to provide access to them. It will also allow us to catalog the websites, make them fully searchable, put them in context, etc. It is a tool not only for preserving websites, but also for providing access to them - as opposed to the Wayback Machine which automatically archives websites. Whereas the Wayback machine is a good resource to look up websites for which the URL is known, an Archive-It collection is a curated collection with websites that are closely monitored by us and which are crawled at frequencies specified by us (another difference to the Wayback Machine).
- Title VI Directors, in an effort to be competitive and innovative with regard to the next round of Title VI applications, are encouraging all groups who receive cooperative funds - such as Outreach directors, language teachers/coordinators, and librarians – to think of projects which support and relate to the priorities outlined in the upcoming call for proposals. While the call has not been issued yet, at a recent Title VI meeting for area directors and associate directors, a senior Title VI program officer pointed out that priorities for the new round of Title VI proposals will be in accordance with President Obama’s stated goals for developing “global competencies”. Specifically, in our context, this means diversity/capacity building for Historically Black Colleges and Universities (HBCUs) and other colleges with smaller or no area studies programs; and access to international information and resources for these colleges. Creating open access web resources, as outlined in this proposal for a web archiving project, will be directly related to these priorities.
For more information, see:

- “Succeeding Globally Through International Education and Engagement: U.S. Department of Education International Strategy 2012-16”, <http://www2.ed.gov/about/inits/ed/international/international-strategy-2012-16.pdf> (4-17-13)
- KimOanh Nguyen-Lam and Cheryl Gibbs, “Measuring the Impact of National Resource Centers: Perspectives from the U.S. Department of Education, International & Foreign Language Education Program, http://easc.osu.edu/sites/easc.osu.edu/files/IFLE_OSU%20Evaluation%20Conference.pdf (4-17-13), presentation at the “NRC Conference: Demonstrating the Impact of National Resource Centers”, in February at Ohio State University, <http://easc.osu.edu/events/2013-nrc-conference>

Proposal for pilot project: Archiving Websites Related to African Countries in Conflict

Over the course of three years, we will archive web sites (including blogs, twitter, podcasts) which focus on countries in conflict, e.g. Mali, Central African Republic, DRC, Libya, et al. This pilot will serve two functions, i.e., the preservation as well as the curation of websites on a specific topic. Many of these sites are temporary, and to preserve access for the future, they need to be captured/crawled while they are active. We do know, for example, that several websites related to Libya have already disappeared or parts of their content has migrated to other websites and/or Facebook pages which are also likely to disappear.¹

¹ Examples provided by Mohamed:

- **Al-Qadhafi talks:** <http://www.algathafi.org/html-english/index.htm> was accessible (last time I used it was April 2011, but sure it was active until the end of al-Qadhafi time with his death). This is what comes now when you try to access it <http://www.algathafi.org/page2cHtoNrQD.php> (a commercial website); however, I believe some of al-Qadhafi followers were able to re-activate the website with some brief contents from the old site (maybe running it from a different country) and here is the link <http://english.algaddafi.org/>
- **Libyan Jamahiriya Broadcasting Corporation** <http://www.ljbc.net/> . This is the link you get now <http://en.ljbc.net/home.php> . When trying to find it using the Arabic name you get this <https://www.ljbc.tv/> which probably was the previous main interface of the disappeared site. Here is the facebook page for the LJBC with some active link to videos and other online resources <https://www.facebook.com/ljbc1> . Here is the current facebook page for the current Libya TV Network <https://www.facebook.com/LTN.TV>

Please also note that these alternative replacement pages including the facebook pages are most probably to disappear soon. As a librarian and a researcher, to agree or disagree with the content of specific material does not mean destroying it. I feel sorry to see these website that contain a long run of materials for more than 40 years of Libya’s history disappeared.

Other sites related to this topic are more permanent (e.g., government websites, news sources), but should be included to present a complete picture of the subject. Our goal is to ensure long-term availability of and access to sites that are of interest to the scholarly community. In the future, we also hope to create a broader network/portal of Africana web resources which would include a wider range of subjects (see section “Beyond the pilot” below).

Budget:

- Archive-It account: \$3000 per year, for three years, to create and maintain the account.
Rationale/discussion item: based Indiana University’s experience with web archiving folklore sites, \$2000 (the budget we originally envisaged) allows for a good number of websites. However, many of these websites are small and infrequently updated. Depending on our expectations, a budget of \$3000 might be more reasonable, particularly if we have sites which need to be crawled frequently.

- Student hourly assistant:
Rationale/discussion item: managing the account, i.e., setting up crawls, monitoring them, solving problems, making adjustments to crawls when necessary is time-consuming work. We are asking for a library school graduate student to assist with this work, for 10 hours per week, at an hourly rate of \$13 per hour. Assuming 40 weeks of work per year, this would add up to \$5600 – which might be too much to handle for the Title VI budget.

Following are two alternative scenarios which try to balance our need for support with our restricted budget:

Both scenarios are based on the assumption that students will not work every week of the year – they will take vacation time, leave for holidays, take time off for exams, etc. There may also be times when one student leaves and another student does not take over immediately. Another assumption is that we may be able to bridge gaps in funding by having interns from the library school.

Scenario I: Amount requested: \$3000

Scenario II: Amount requested: \$4000

In both scenarios, we would use the allocated amount until it is used and then try to bridge the gaps. In Scenario I, this may mean that we will have to come back to the Title VI Librarians group with a request for supplemental funding. Scenario II makes such a request less likely.

Logistics:

- Indiana University Libraries will subscribe to the Archive-It service on behalf of the Title VI Librarians. IU will create an account specifically for the Title VI Librarians, and set up workflow, authorizations and policies per directions from the project editorial board. Marion Frank-Wilson will function as the liaison between IU and the Title VI Librarians.
- The members of this task force (Karen, Tim, Jason, Lauris, Mohamed, Marion) will serve as the project's editorial board. In addition to Marion, Jason and Karen will have access/authorization to directly edit and set up crawls from within the account.
- The websites below have been selected by the members of this task force as the initial websites to be archived.
- Other websites will be added over the course of the next several months. Suggestions for additional websites will be solicited from ALC members, to be submitted to the editorial board.
- With the help of an IU library school graduate student assistant, members of the editorial board will monitor the Archive-It account and budget, determine and adjust crawl frequencies, keep track of broken links and other changes to the websites.
- Other support: based on IU's experience with Archive-It projects, we know that we can rely on Archive-It staff for robust customer support, and for help tailoring the budget. IU Libraries will provide supplemental/advisory support, including workflows, practices, and cataloging advice.

Policy/best practices:

Quite a few institutions are involved with various aspects of web archiving (using Archive-It). Many of these projects relate to university archives while others are subject based projects, as for example Columbia University's Human Rights portal, or our proposed pilot project. Although in practice for several years, the field of web archiving is still emerging and, to our knowledge, there is no commonly accepted set of rules or best practices.

In 2012, Columbia University hosted the "Web Archiving Policies and Practice in the US: 2012 Summit", with "invited participants from over 20 US academic and research libraries to share perspectives on topics relevant to their own experience". Among other topics, participants of the summit discussed policies regarding permissions and control of content, and documents are posted on the summit's website at <https://webarch.cul.columbia.edu/>. Within the first six months of the pilot project, the editorial board should review these documents and develop a detailed policy for our pilot project.

More immediately, to get us started: as a result of an extensive web collecting/archiving program, Columbia University has posted a policy which has, for example, been adopted and modified for Indiana University's Open Folklore site. We propose, for the purpose of launching our pilot project, to adopt Columbia University's policy regarding permissions:

"We will attempt to obtain permission to archive from the website owner before harvesting content, and respect any denial of permission. If no response is received to an initial request, we will send a follow-up reminder, announcing our intent to archive if the site owner does not explicitly object.

Permission to continue archiving will be assumed to remain in effect indefinitely unless otherwise stated or explicitly revoked. If the organization granting permission changes name but the web address remains the same, or if the website address changes but the organization name remains the same, no new permission will be sought. If both the organization name and website address change, the new site will be considered a separate organization and new permission will be sought.

If a site's content has been harvested and archived without explicit permission (e.g., no response was received to our request or notification), we will honor requests to cease archiving and, to the extent possible, to remove content from the archive.

Once permission to archive has been granted, we will remove content only in exceptional circumstances. These include:

- a) official notification that the individual who granted permission did not have the right to do so;
- b) notification that the site owner had no authority to grant permission to archive specific content with a different rightsholder;
- c) notification that specific content has been removed from the site because of a legal challenge, such as a claim of libel.

When content that is stored with an external agent (such as Internet Archive) must be removed, we will act to the best of our ability to ensure that this happens. At minimum, we will block public access to the content (even if that requires blocking access to additional content from the same site.) “

Addition from the task force: There might be cases when a name or some content crawled and open to the public on the web has endangered or could endanger someone. Even if someone in authority originally gave permission, we should be willing to remove content which might endanger someone.

Initial set of websites to be archived:

For Mali:

- Le Republicain. Mali newspaper. <http://lerepublicain-mali.com/>
This might be crawled as an example of a newspaper site. One of their journalists was jailed for what appeared in the newspaper; he has been released. Article about it – <http://www.cpj.org/2013/04/malian-journalist-boukary-daou-released-on-bail.php#more>
- Mouvement National de Libération de l'Azawad, MPLA. <http://www.mnlamov.net/>
(another, similar, site has already disappeared)
- Bamako Culture - <http://www.bamako-culture.org/>
Was an online (maybe print ?) magazine. Seems to be no longer active and might disappear.
- Bridges from Bamako. <http://bridgesfrombamako.com/>

Blog (Wordpress) by Bruce Whitehouse, a former Peace Corps Volunteer residing in Bamako, Mali. Has posts this month. It has an analysis of the 2012 political crisis By Baz Lecocq, Gregory Mann, Bruce Whitehouse, Dida Badi, Lotte Pelckmans, Nadia Belalimat, Bruce Hall, and Wolfram Lacher.

- MaliWatch. <http://www.maliwatch.org/>
An non-political NGO originated in Washington DC. Had projects, not especially active lately, but an example of work by the Diaspora. They might like to have their work archived.
- Mouvement National de Libération de l'Azawad, MPLA. <http://www.mnlamov.net/>
In French. Tuareg rebel group opposed to the Mali government. Declared the northern Mali region they captured independent, April 2012.

For Central African Republic (recommended by Natalia Taylor Bowdoin, ALC member, former Peace Corps volunteer in CAR):

- <http://www.radiodekeluka.org/>
- <http://www.centrafrique-presse.info/site/>
- <http://www.centrafriquelibre.info/>
- <http://centrafrique-presse.over-blog.com/>
- <http://reseaudesjournalistesrca.wordpress.com/>

For DRC:

- Also recommended by Natalia: <http://www.renadoc.org/?p=449>

Recommended by Simon Bockie:

- Media Congo:
<http://www.mediacongo.net>
- Infobascongo:
<http://www.infobascongo.net>
- Radio Okapi:
<http://radiookapi.net>
- Kleber - L'Observateur:
<http://kleber-observateur.blogspot.com>
- Congo Siasa:
<http://congosiassa.blogspot.com>

- Soiete Civile.cd:
<http://www.societecivile.cd>
- Churches, Peacebuilding and Women's Rights in DRC:
<http://www.drcongo.nibrinternational.no>

Selection criteria:

Since it is unpredictable which websites will remain and which will disappear, we chose not to use 'danger of disappearing' as our main criterion for selection. Instead, for this pilot project we decided on the subject specific approach of selecting websites related to African countries in conflict. We believe that this approach will necessarily include websites in danger of disappearing (see example of ceased Libya websites mentioned above).

We chose this subject/topic because of its importance and relevance to the scholarly community: as we have seen in Mali, CAR, Libya et al., as civil unrest occurs, information is most immediately disseminated on websites, blogs, and other electronic discussion fora. This information – which often can be considered primary source material - is of research interest and needs to be preserved before it vanishes. In addition to the preservation aspect, building this web archive will also allow us to curate, and provide access to a collection of websites on a topic of current interest – which may well be of interest to researchers in the future..

Our objective is not to be comprehensive, but to collect material that is either of current interest to scholars or likely to be of interest for future research.

Beyond the pilot:

This pilot project will help us explore the feasibility of creating a more in-depth portal for Africana websites, extending into many more subjects. Such a portal would also include links to already existing web archives. While there are web archiving projects with Africa-related content, such as Columbia University's Human Rights Portal or the African Political Web Archive maintained by Karen Fung, in our research to date we did not find a portal specifically for Africana sites across a range of subjects.

Questions this pilot project will help us answer:

- Logistics: this project will help us develop an efficient workflow to manage a web archiving project.
- Also related to logistics, we will learn how to archive different types of websites (such as twitter and blog posts, podcasts, etc.).

- We will gather experience with the selection process; e.g., is our idea to funnel website suggestions to the editorial board for consideration feasible? Or should we reach out to the larger scholarly community and develop a scholar-driven selection process?
- Best ways to promote this project among faculty, graduate students, researchers.
- Ultimately, the goal is to conceive of a larger project, a portal to archive Africana resources.
- Based on our experience with the pilot project, will we be interested in pursuing the idea of a portal? If yes, would we use the same approach of an ALC editorial board? Would we even be able to create and sustain such a portal without partners? Or would we look for partners and collaborate, for example with the African Studies Association?
- We will gain insight into the cost of such projects, and might begin exploring sustainable funding models.