

# FOCUS *on Global Resources*

Fall 2011 • Volume 31 • Number 1

## MIDDLE EASTERN and ISLAMIC RESOURCES

- Provincial Turkish  
Newspapers and the  
Negotiation of a Muslim  
National Identity 2
- Documenting Revolution  
in the Middle East 5
- An Imperfect History:  
Capturing the Middle  
East Web 8
- Middle Eastern and  
Islamic Resources at CRL 13

The periodical *Soera Moehammadijah* documents the development and programs of Muhammadiyah, Indonesia's largest and most successful modernist Muslim movement. CRL collections.

[www.crl.edu](http://www.crl.edu)



### *In This Issue*

This issue of *Focus* explores, from a number of perspectives, the variety of source materials that can be brought to bear on scholarly research on the Middle East. James Simon discusses a new book by Gavin D. Brockett of Wilfrid Laurier University, which traces the role played by provincial newspapers in shaping a Turkish national identity after the end of the Ottoman Empire. Professor Brockett's book relies heavily on newspapers from the CRL collections for a street-level view of this process. "Middle Eastern and Islamic Collections at CRL" features some newly acquired collections in this area, from Arabic manuscripts from the British Library and the University of London to British intelligence files on Iraq.

For insights on the momentous recent events in Egypt, Libya, and Tunisia, historians and public policy researchers will employ new forms of evidence. In "Documenting Revolution in the Middle East," Roberta Dougherty of the University of Texas surveys the myriad types of records generated by the recent Arab Spring, including web communications and postings on social media. James Simon's careful analysis of the various efforts to archive ephemeral Web communications identifies the advances, and the limitations, of current best practices in this arena.

—Bernard F. Reilly, Jr.  
President

# Provincial Turkish Newspapers and the Negotiation of a Muslim National Identity

James T. Simon

Director of International Resources  
Center for Research Libraries



The Turkish newspaper *Demokrat Hamle* (August 30, 1953) is one of the many CRL resources Professor Gavin D. Brockett used to research his new book, *How Happy to Call Oneself a Turk: Provincial Newspapers and the Negotiation of a Muslim National Identity* (University of Texas).

In 1923, the Republic of Turkey was born from the ashes of the Ottoman Empire, following the tumult of World War I and the conflicts engulfing Anatolia and Eastern Thrace through the Turkish War of Independence. The leading figure in the Turkish nationalist movement was Mustafa Kemal Atatürk, founder of the Republic and by most accounts the key force in transforming Turkey into a modern secularist state.

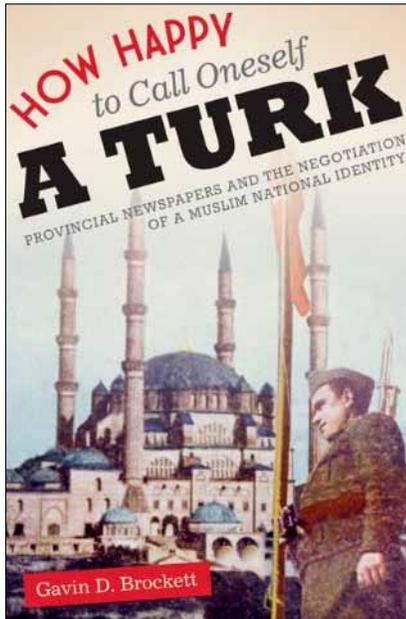
The bulk of the scholarly record credits Mustafa Kemal Atatürk with creating a Turkish national identity through his progressive secularization of the country during his presidency (1923–38). However, in his new publication, Gavin D. Brockett, Associate Professor of Middle East and Islamic History at Wilfrid Laurier University, challenges the dominant narrative of a united Turkish national identity under Atatürk by examining the press history of the nation, particularly regional newspapers held by CRL and other libraries.

*How Happy to Call Oneself a Turk: Provincial Newspapers and the Negotiation of a Muslim National Identity* (University of Texas Press, 2011) traces the development of the “national” print culture of Turkey from the founding of the state through the transition to a multiparty democracy in the 1940s and ’50s. Brockett focuses on the press outside metropolitan centers to understand how the principles of secularism (*laiklik*) and modernization were perceived beyond the Kemalist elite. In analyzing how print contributed to the formation of a common national identity, Brockett reveals far more complexity than the traditional political narrative has previously conveyed.

## “Serendipitous Find”

As a doctoral student at the University of Chicago, Brockett never intended to embark on this particular research pathway. In searching for articles by a particular Turkish intellectual known to have circulated widely, Brockett came across a listing of CRL’s extensive holdings of Turkish newspapers. Borne more of curiosity (“professional procrastination”) than direct research application, Brockett borrowed a selection of papers through interlibrary loan. He marveled at the variety of news he discovered in these remote papers: images from abroad such as President Truman throwing the ceremonial first baseball pitch in 1951, and photos and political cartoons clipped from foreign media. As he read further, however, he began to note the diversity of local geographic, political, and religious perspectives promoted in the various papers. “To a researcher, opening up those wonderful brown paper packets,” Brockett notes, “was like striking gold.”

Eventually abandoning his original research, Brockett started ordering more papers



In his book, Brockett used newspapers archived at CRL to examine social and political debates from around Turkey. He maintains that the unprecedented expansion of print media in Turkey between 1945 and 1954 created a forum in which ordinary people could inject popular religious identities into the new Turkish nationalism.

from CRL, which were delivered on large trolleys. Brockett received space at the university library to store and use the papers. After three years of research, he relates, he still had not exhausted the collection.

### Completing the Research

After Brockett completed his doctorate at Chicago, he spent a number of years on research to flesh out his work into a full publication: “When I first started reading the papers from CRL, I found there was no comprehensive catalog or index for these papers.” Gleaning information from diplomatic reports and press summaries was helpful in identifying additional resources. But the most important information came from the papers themselves. Newspapers often printed information about other publications, in some cases pricing and circulation figures. References to competitors (frequently disparaging) provided Brockett a list of other targets to pursue, but it required additional research to track down titles, particularly for the earlier years. “The reality of the collection [at CRL] is that it is not complete. CRL holds a lot of the known papers—the majority of them—but coverage begins largely in 1950, and ends around 1953.” Receiving research grants, he ventured into collections in Ankara and Istanbul to supplement the materials found at CRL.

Press laws in Turkey ensured that copies of all published works (including newspapers) were deposited in several locations. The National Library of Turkey (*Millî Kütüphane*), established in 1950, possesses the most extensive collection of newspapers in the country. Brockett spent time poring through the collections in the limited time he had available. “There are countless challenges to accessing these collections. Cataloging is not always reliable. There can be bureaucratic red tape, such as actually gaining access to a library or only being allowed to see a limited number of titles each day. Fortunately, the staff began to tire of my repeated requests to haul these heavy volumes and eventually allowed me into the depot to pursue my research. After this I was able to be much more productive”.

“There is no question the research in Turkey assisted in my work,” Brockett notes. “However, it did make me value the comparative ease of access to CRL’s collection.” He also appreciated the assurance that a second copy of the material was accessible should misfortune befall the originals.

He adds that additional material exists in private collections in the region, but trying to identify those collections and gaining access to them is quite difficult. In one case, after identifying a title referenced in another paper held by CRL, he engaged in an extensive search in Turkey. “Time and again I was told the paper did not exist. Finally, I went to Samsun [on the coast of the Black Sea] and found the editor, still alive in his late seventies. He had about half of the run, and he generously photocopied his holdings for me.” Brockett eventually found the rest, uncataloged, while working in the newspaper depot in Ankara.

### Challenging the National Identity

Brockett’s work challenges the assumption that the transformation of the country and formation of a national identity can be told solely from the perspective of Atatürk: “It’s like telling the story of America from the perspective of George Washington, ignoring everyone else.” Moreover, the Kemalist vision of a secularist society ignores the significant role that religion played—and continues to play—in Turkish identity. Brockett argues that the implementation of *laiklik* actually proved to be an obstacle to national identification for those Turks for whom religious identities remained important. It was not until the formation of a multiparty state, and the

resulting expansion of the press, that the people of Turkey could properly explore the relationship between national and religious identities.

The evidence, found in the newspapers, shows that when Turks had the opportunity to express themselves in print, topics frequently revolved around religion. Religious newspapers were very important in the scheme of the provincial press. These papers ranged in scope and purpose, from the strictly religious (publication of sermons, articles about proper beliefs and practices as Muslims) to those engaged in the political debates of the day (regulating forms of religious practice or the language used for the call to prayer). At their core, the provincial press—largely ignored in the historical record, according to Brockett—provided a direct challenge by the surrounding populace to the centrally imposed principles of *laiklik*.

Not to say these religiously oriented papers were anti-nationalist. According to Brockett, nearly all of them were very pronationalist. However, the tenet of their argument was that a “good Turk” was also a Muslim Turk, and that every Turk had a right to be proud of being a Muslim. This negotiation between the nationalist ideology of Atatürk and the traditional principles held by the population is central to understanding the identity that Turkey has today.

### Future Research

Brockett notes that much fertile research lies ahead. The ongoing identification and study of these under-represented resources, as well as surviving personal archives about the publications or publishers, serve as grist for future scholars. Also of significance, he relates, would be the records of the government’s suppression and closure of newspapers. These records, along with information on trials and persecution of publishers and reporters, remain closed to the public.

Brockett advises libraries to work closely with scholars who have intimate knowledge of collections and the availability of potential resources. He notes, for instance, the presence of a vibrant online second-hand book market in Turkey, in which many rare publications are offered. He cautions that these resources are often bought by private collectors, and then disappear from view. While he acknowledges that all libraries have finite resources, Brockett encourages institutions to think broadly about how to work with researchers to shore up the historical record and make these lesser-known but significant resources available to future scholars. ❖

## Documenting Revolution in the Middle East

**Roberta L. Dougherty**  
Middle Eastern Studies Librarian  
University of Texas at Austin



This photograph of a Tahrir protest was taken on September 9, 2011, and posted in Kodak Agfa's photostream on Flickr. Courtesy of Flickr.

What does it mean to document a revolution? The recent massive uprisings in Arab countries created a plethora of artifacts: printed documents of various kinds used by protesters to organize their efforts and codify their ideals, demands, and goals; photographs and video recordings of protesters in the street, news broadcasts, televised speeches by beleaguered leaders; posters and signs used to motivate the efforts of both revolutionaries and counter-revolutionaries; creative responses in the form of art, music, multimedia performances; and websites and social media tools for organizing, motivation, and documentation.

Documentation of the so-called “Arab Spring” involves collecting a number of formats, some born-digital, some analog, and all almost entirely ephemeral, fleeting, and endangered.

Some significant projects have emerged that attempt to document the revolutionary movements that began in December 2010, particularly Egypt's Tahrir Square:

### **The Committee to Document the 25th January Revolution—**

This volunteer group consists of historians, political scientists, anthropologists, and technology experts “collecting everything we know, including oral testimonies, blogs, newspaper clippings, even Facebook status messages and tweets in order to make this a wide library resource for anyone studying the events.”<sup>1</sup> The group organized on January 14, 2011, the date of Tunisian president Ben Ali's resignation.

Khaled Fahmy, one of the country's leading historians, was asked by the head of Egypt's archives to oversee the project to document the country's political and social upheaval, “to gather as much primary data on the revolution as possible and deposit it in the archives so that Egyptians now and in the future can construct their own narratives about this pivotal period.”<sup>2</sup> As Fahmy described to *The Guardian*: “Documenting the revolution sounded like an easy thing, but what is the revolution? When did it start? When did it end? What constitutes participation in the revolution—is it only those who went down to Tahrir, or is it also the doctors who worked extra-long hours in their hospitals to treat the wounded? What about a police officer who fought the protesters—is he a part of the revolution or not?”

**R-Shief**—R-Shief is a data-mining project that draws content from Twitter and various websites and blog posts. It attempts not only to document the Arab Spring, but also provides tools for visual analysis of this data. Its “Twitterminer” tool (one of the very few places that provided public access to tweets posted

<sup>1</sup> Mary Mourad, “We Will Document Egypt's Revolution, not Dictate the Story: AUC Professor,” *Ahram Online* May 28, 2011, accessed September 30, 2011, <http://english.ahram.org.eg/News/13026.aspx>.

<sup>2</sup> Jack Schenker, “The Struggle to Document Egypt's Revolution,” *The Guardian* July 15, 2011, accessed September 30, 2011, <http://www.guardian.co.uk/world/2011/jul/15/struggle-to-document-egypt-revolution>.

during the Arab Spring), permits data analysis using preselected hashtags. Media artist and critic Laila ShereenSakr (who also publishes under the name “VJ Um Amel) founded this project in 2008.

**“Arab Spring: An Interactive Timeline of Middle East Protests”**—

Developed by the UK’s *Guardian* newspaper, this beautifully designed tool provides elegant access to the web-based archive of *Guardian* news stories on the “Arab Spring.” The timeline begins on December 19, 2010, with the news of the December 17 self-immolation of Tunisian Mohammed Bouazizi, and has been continuously updated (last viewed September 7, 2011).

**“Collection: North Africa and the Middle East 2011”**—The Library of Congress has established several projects to archive the web in partnership with the Internet Archive, using IA’s subscription service Archive-It in collaboration with other research and memory institutions around the world.

LC’s web archive for North Africa and the Middle East (collaborating with Archive-It’s Global Events project and the National Library of France) began in February 2011, although the collection also includes related documents (blog posts, etc.) that are a few years older. LC also did a separate project in partnership with the same two institutions devoted to archiving web content related specifically to the revolution in Tunisia. [CRL’s James Simon discusses this project further on pp. 8–12.]

**AUC’s “University on the Square”**—With financial support from the Andrew W. Mellon Foundation, this project was initiated in the spring of 2011 at the American University in Cairo (AUC) to take advantage of the university’s unique position, with its downtown campus located right on Tahrir Square. In addition to collection posters and artwork, the project included gathering oral history from students and others affiliated with the university about their experiences during the spring uprisings. The website created as part of this project includes an image gallery and links to video files contributed by AUC students and faculty.

**Tahrir Documents**—This initiative focuses on documenting the printed ephemera of the spring 2011 uprisings in Egypt. The website provides scans of dozens of printed leaflets, from religious tracts to lists of political demands and handbooks for protesters. It is particularly useful for providing not only scans of the original documents, but also translations into English, done by a team of volunteers. The website first began posting in March 2011.

**25Leaks.com**—The individual initiative 25Leaks.com, which also began in March 2011, is an ambitious attempt to document and provide access to materials seized by protesters from the infamous state security headquarters after Mubarak was ousted from office. The site is in Arabic only, and the site’s creators have remained anonymous for their own safety.

## Documentation Using Social Media

Among several attempts to use Facebook as an archive is “[Revolution Graffiti](#),” which created a Facebook page on March 4, 2011. The page description reports: “This page is mostly to gather all revolution graffiti images in one place. In other words archiving the graffiti. By revolution graffiti I mean graffiti created from the start of the revolution till now and later on.” The page offers a variety of artwork, projects, articles, and comments related to uprisings in regions like Cairo and Libya.

The Facebook page of “[I Am Tahrir: The Art of Revolution](#)” offers an online “exhibit that collects art produced during or inspired by the January 25 revolution—and

not just in digital format. The initiative aims to document the days of the revolution, through art, but its vision extends beyond this scope. We'd like to be a hub for "revolutionary art" from across the world—a place where the hands of art censors and staunch conservatives cannot reach."

Many photographers also uploaded revolution-related photographs to flickr. The biggest problem with this kind of resource is self-curation: specific people could be tagged but are not; precise metadata is often not available; and photos might be tagged or not at all.

In previous collaborative projects, the Library of Congress and the American University used Archive-It to collect various kinds of web content, including Twitter feeds. However, it remains difficult to "excavate" relevant content from archived Twitter feeds using hashtags.

The new publication *Tweets from Tahrir: Egypt's Revolution as it Unfolded, in the Words of the People Who Made It* (OR Books), edited by Alex Nunns and Nadia Idle, paints "an exhilarating picture of an uprising in real-time," containing Twitter updates by thousands of young people who documented on cell phones every stage of the revolution, as it happened. *The New York Times* called the book "a feat of nearly real-time publishing." It demonstrates how technology has undoubtedly changed the way that wars, battles, uprisings, and revolutions will be documented from now on. ❖

# An Imperfect History: Capturing the Middle East Web

James T. Simon

Director of International Resources  
Center for Research Libraries



This screenshot of the online publication *al-akhbar* shows Egyptians looking at documents within the State Security Investigations Service offices in Cairo. Courtesy of Archive-It.

In 2005, as part of its Political Communications Web Archiving investigation, CRL conducted a pilot assessment of Archive-It, a subscription service developed by the Internet Archive to harvest, catalog, manage, and display web-based information. Archive-It allows subscribers to select born-digital content for harvesting on a flexible crawling schedule (from daily crawls to annual, in addition to “on demand”).

CRL issued a report in 2006 on “[Middle Eastern Political Parties Web Harvesting and Other Efforts](#),” assessing the crawling capability and administrative tools of Archive-It in the context of a subject-based harvest of websites of Middle Eastern political parties. The detailed report highlighted successes as well as deficiencies in capture, and discussed broader concerns over the complexity of managing and making discoverable archived content.

Five years later, web archiving continues to pose significant challenges to libraries and archives around the world. National Libraries and cultural heritage institutions have coalesced around the International Internet Preservation Consortium (IIPC) to foster the development and use of common tools, techniques, and standards that enable the creation of international archives. The Internet Archive plays a key role in this initiative, providing the open source crawler technology based on IA’s Heritrix web crawler for further development.

Meanwhile, Archive-It has expanded its subscriber base to more than 175 academic, government, and nonprofit partners, reportedly harvesting more than 3.2 billion URLs contained in 1,687 public collections (as of September 2011). In this sea of content, numerous collections have developed around Middle Eastern and Islamic themes. These include one-off capture of organizations’ sites (IslamAmerica, Institute for Palestine Studies) to event-based captures of such world events as the Libyan uprising and the Tunisian revolution.

To evaluate how such crawling technologies currently serve research purposes, CRL recently performed an assessment of Archive-It content based on three public collections related to the Middle East. As the samples below demonstrate, a combination of human and technical pitfalls may cause errors in capture, resulting in an “imperfect history” of these events.

## CRL Assessment of Archived Web Collections

Iranian Blogs

<http://www.archive-it.org/collections/1035>

<http://www.archive-it.org/collections/1147>

<http://www.archive-it.org/collections/1372>

<http://www.archive-it.org/collections/1749>

Partner: [New York University, Islamic and Middle Eastern Collection](#)

Crawling Activity: 2008–2011

Sites crawled: approximately 755

This collection of sites (split among multiple collections within Archive-It) features an array of blogs produced by Iranian scholars, politicians, journalists, and the general public. Blogging factors prominently in Iranian youth and opposition movements. While print and broadcast media is tightly controlled by the state, a large percentage of Iran's young population turns to the Internet and social media for free expression. During the parliamentary election in 2008 and more prominently following the presidential election in 2009, opposition sites were routinely blocked by authorities and many bloggers and journalists were suppressed and detained. Some of the collection's sites are still active on the web, but many are no longer updated or have since been removed entirely.

As blogs tend to feature less-advanced programming and follow a consistent template, content capture for these sites is reasonably successful. Embedded videos frequently do not work, or else point to live versions of videos hosted by external sites (e.g., YouTube, *Al Jazeera*). External links were generally not included in the crawl, limiting the functionality of posts that refer to external news items or posts (a common occurrence).

The archived collection includes a number of prominent sites, such as that of Mohammad Ali Abtahi<sup>1</sup>, former Vice President and reformist arrested in 2009; and Hossein Derakhshan (aka Hoder),<sup>2</sup> widely considered the “father of Persian blogging,” who was arrested on charges of spying for Israel. Missing from the collection are sites of several other well-publicized bloggers such as that of Mohammad Pour Abdullah,<sup>3</sup> a student activist sentenced to six years in prison on charges of anti-government publicity; or Omid Reza Mir Sayafi,<sup>4</sup> who was similarly detained and later died in prison.

Across the various collections archived by NYU, roughly 755 unique seed URLs were input for crawling (not all submitted sites returned crawl results). The periods of capture vary according to each phase of crawling, but span from early 2008 to the present. Many sites were crawled only once or for a brief period of time, with some pages crawled more intensively over longer periods. Information about selection and frequency methodologies is not publicly available.

The Iranian blogosphere is large (by varying accounts, some 60,000 to 100,000 blogs are updated regularly). Given this scale, the Archive-It collection can be considered representative at best. For the Berkman Center for Internet & Society's 2008 study “*Mapping Iran's Online Public: Politics and Culture in the Persian Blogosphere*,”<sup>5</sup> Morningside Analytics tracked over 200,000 Persian language blogs, including 98,875 blogs monitored daily. While the Berkman Center harvest focused on text analysis rather than full-page rendition, it is evident that a combined approach using social network analysis and robust web-crawling technology could feasibly harvest a more complete picture of the blogosphere.

### 2011 Egyptian Revolution

<http://www.archive-it.org/public/collection.html?id=2358>

Partner: [American University in Cairo](#)

Crawling Activity: 2011

Sites crawled: 90

<sup>1</sup> <http://www.webneveshtehe.com>, accessed September 20, 2011. Archive-It page at [http://wayback.archive-it.org/1749/\\*http://webneveshtehe.com/](http://wayback.archive-it.org/1749/*http://webneveshtehe.com/)

<sup>2</sup> Archive-It page at [http://wayback.archive-it.org/1035/\\*http://i.hoder.com/](http://wayback.archive-it.org/1035/*http://i.hoder.com/)

<sup>3</sup> <http://www.feuer17.blogspot.com/>, accessed September 20, 2011.

<sup>4</sup> <http://rooznegaar.blogfa.com>, no longer accessible. Press at the time pointed to an Internet Archive capture of this site, but this collection is no longer viewable via the Wayback Machine (the archived page now reads “Page cannot be crawled or displayed due to robots.txt”).

<sup>5</sup> [http://cyber.law.harvard.edu/publications/2008/Mapping\\_Irans\\_Online\\_Public](http://cyber.law.harvard.edu/publications/2008/Mapping_Irans_Online_Public), accessed September 20, 2011.

This collection of crawled sites relates to the popular uprising in Egypt that began on January 25, 2011. The collection, curated by the American University in Cairo, began its crawl on February 1, capturing events leading up to the resignation of Hosni Mubarak and the ongoing transformation of Egyptian government and society.

According to the Archive-It collection notes, the following categories of sites were crawled:

- Blogs and Twitter Feeds (25 URLs)
- Documentary Projects (5 URLs)
- Memorial Websites (3 URLs)
- News and Media Coverage (40 URLs)
- Photos and Videos (4 URLs)
- Related Websites (13 URLs)

An analysis of the harvested sites reveals significant problems in capture for many. While blog sites tend to be more easily crawled, often harvested back to the earliest posts of individual bloggers, external links are frequently “not in archive.” Embedded streaming videos frequently were not captured. The links to videos in most cases produced errors or pulled from videos hosted on external sites where still available. Crawls of Twitter pages often did not capture any links or information beyond the first visible screen.<sup>6</sup> Particularly problematic for capture were the documentary and “memorial” websites (such as [iwasintahrir.com](http://iwasintahrir.com), [iamjan25.com](http://iamjan25.com), and [18daysinegypt.com](http://18daysinegypt.com)). These pages draw heavily on JavaScript functionality and either were not properly captured or could not be rendered and viewed—with or without JavaScript enabled.

News and media coverage varies significantly in success rate. News and breaking event pages frequently captured headline text and related articles, but photos, videos, and style sheets needed to properly render the pages as displayed frequently were not captured. In some cases, related sections of news sites were included; in others they were not captured.

A small selection of Wikipedia pages on figures such as Hosni Mubarak and Mohamed ElBaradei shows how articles describing current events undergo rapid and successive changes over time. It should be noted, however, that Wikipedia itself maintains revision histories of its pages at a more granular level than a periodic web harvester can capture.<sup>7</sup>

With a seed list of 90 URLs, this collection represents a small sampling of potentially relevant sites, though the inclusion of particular sites over others leads to questions of selection and objectivity. Contextual information on page selection would be particularly useful in this type of collection.

## North Africa and the Middle East 2011

<http://www.archive-it.org/collections/2349>

Partner: [Internet Archive Global Events](#)

Crawling Activity: 2011

Sites crawled: 5000+

Collected by the “Internet Archive Global Events” effort, this initiative sponsored by the Internet Archive is responsible for capturing event-based collections such as the [Jasmine Revolution – Tunisia 2011](#) and the [earthquake in Haiti](#). According to the Archive-It site, this collection includes blogs, social media, and news sites about Egypt, Yemen, Libya, Sudan, and other countries documenting the tumultuous events in Northern Africa and the Middle East starting in January 2011. With an impressive 5,178 seed URLs crawled, this collection appears to be significant, with

---

<sup>6</sup> The Archive-It “[Help Wiki](#)” indicates earlier crawls required an “expanded scope” to capture additional content beyond the “more” button. In addition, the wiki suggests that in some cases, disabling JavaScript in the viewing browser is required to view content. This was not successful during the period of review.

<sup>7</sup> See, for instance, [http://en.wikipedia.org/w/index.php?title=Mohamed\\_ElBaradei&action=history](http://en.wikipedia.org/w/index.php?title=Mohamed_ElBaradei&action=history) versus the Archive-It crawl at: [http://wayback.archive-it.org/2358/\\*http://en.wikipedia.org/wiki/Mohamed\\_ElBaradei](http://wayback.archive-it.org/2358/*http://en.wikipedia.org/wiki/Mohamed_ElBaradei), accessed September 20, 2011.

content reportedly contributed by partners at the Library of Congress, Bibliothèque nationale de France, the British Library, and Stanford University.

Of the crawled URLs, 4,741 (92%) point to YouTube channels or individual videos, many of which were not properly configured for capture. With Archive-It, videos hosted by YouTube are best crawled one-by-one, and cannot be viewed in the archived page. Documentation on how to access the archived video is not easily discoverable, though Archive-It is currently testing implementation of a linked video page for viewing.

Of the 437 remaining sites, 133 (30%) were individual or organizational Facebook pages, 71 (16%) were readily identified as blog posts, 49 (11%) linked to Twitter feeds, and 35 (8%) linked to not-for-profit sites. The remainder is a mix of news organization pages and portals, memory sites (<http://www.iamjan25.com/>, <http://1000memories.com/egypt/>), video collections, discussion boards, and official sites of journalists and public figures.

As with other collections, many of the crawls suffered capture problems, particularly video links and Java-based functionality. News links may capture top-level aggregated pages, but links to detailed articles are often not available.

While the collection description includes the topic of “Government,” notably absent from the collection is the presence of government pages or sites that represent the perspective of official regimes. With the exception of the page for the Government of Southern Sudan (<http://www.goss.org/>), no other government sites were included.

A small amount of overlap occurs between sites selected for this collection and for the “2011 Egyptian Revolution” collection (above). This occurs most frequently with news organization sites, though the period and frequency of crawls in this collection appears much more extensive than the American University in Cairo collection.

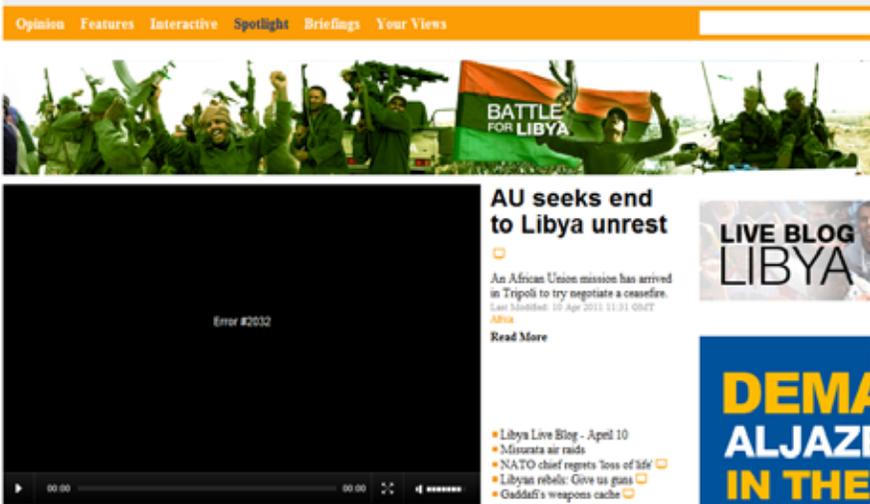
## Conclusions

From the research perspective, given the numerous limitations described above, it is difficult to draw firm conclusions on the utility of the archives for historical research. Without these efforts, many of the sites would not be available today in any form (much less the reduced form in which they currently appear). The recommendations below are intended to guide library partners in future web-archiving efforts.

**Curatorial Challenges:** The selection of sites and determination of crawl frequency remains one of the more time-consuming and challenging aspects of web archiving. The number of sites selected that were essentially “uncrawlable” is quite substantial in the case studies above. Seed URLs that point to too wide a capture scope (the entire site of the Bibliotheca Alexandrina appears to have been crawled in its entirety multiple times for the Egyptian Revolution archive) or too narrow (a single page of a multiple page PDF document) create inconsistencies in the collection focus. Selectors must be further educated in appropriate URL identification and have a priori knowledge of what types of sites cannot be adequately captured.

Providing contextualized access to the archived collections offers another curatorial challenge. For the three case studies above, there did not appear to be any guides that indicate the scope of the collection, content selection criteria, or links to particularly notable sites or pages. While Archive-It has improved its search functionality (including the ability to search in vernacular scripts) and provides advanced search capabilities, navigation and use of collections in a cross-temporal archive remains difficult.

**Administrative Challenges:** Checking the quality of capture and appropriateness of the frequency is another consistent challenge in web-archive collections.



Some of Archive-It's online pages fail to capture live elements such as video, as shown here by the black error screen on this Libyan site. Courtesy of Archive-It.

Certain sites that had not changed in months (or years) were still included in numerous crawl schedules. While Archive-It recommends reviewing crawl reports and browsing archived documents, such quality review appears to be undertaken with varying vigor by particular partners.

**Technical Challenges:** Many of the technical challenges in harvesting are well-known in the web-archiving community. Archive-It provides documentation and tips for its partners and Help Wiki, and is diligently pursuing technical solutions in response to developments in technology or changes in access to categories of sites.

Perhaps the biggest challenge of web archiving is not only to capture the content of a site but the full experience of the web at a given point of time. How can libraries and archives reconstruct the experience so that future researchers of the 2011 Egyptian Revolution can study what was happening on the web—the entire web—during the period leading up to and following the events? The Internet Archive is exploring this scenario through the beta “replay version” of the Wayback Machine.<sup>8</sup>

Libraries need to explore further how researchers, policy makers, and the commercial sector are using web archives, and what types of resources they require. A 2011 report from the Oxford Internet Institute entitled “Web Archives: The Future(s)” suggests that existing web-archive efforts have not seen wide takeup by the research community. Rather, researchers increasingly view the live web as the archive, with data loss “outweighed for the most part by the otherwise huge volume of data that remains on the web at any given time.”<sup>9</sup>

Understanding the user needs can inform libraries efforts in archiving content and the layering of tools on top of archives. The Oxford report recommends that the “web archiving community needs to connect the resources they are building with the cutting edge tools being developed by computer scientists, researchers, independent developers, and hackers to study the live web.” Social Network Analysis, data visualization, APIs, and linked data are all means of exploring the live web that, if applied to archived web data, can assist researchers’ quest for understanding and interpreting our own imperfect history. ❖

<sup>8</sup> The beta version of the Wayback Machine features calendar to browse a site on a particular date, from which you may be able to visit linked sites crawled on or about the same time period. See <http://www.archive.org/web/web.php>, accessed September 20, 2011.

<sup>9</sup> Meyer, Eric T., et al, “Web Archives: The Future(s),” [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1830025](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1830025), accessed September 20, 2011.

## Middle Eastern and Islamic Resources at CRL



Graphic from a January 1941 edition of the Indonesian Muslim periodical *Soera Moehammadijah*. CRL collections.

Since the last report on Middle Eastern resources at CRL (see: *FOCUS*, Spring 2004, Volume 23, Number 3), CRL has substantially augmented its collections in response to member feedback. Acquired through CRL's collection-building programs, demand purchase, or cooperative programs, the resources described below represent critical primary source collections of interest to scholars today. For a more in-depth review of CRL's collections and services in this area, view the presentation of the August 2011 [Webinar on Middle East and Islamic Resources](#).

### Arabic Manuscripts in the British Library

The British Library's collection of Arabic manuscripts numbers over 20,000 works, one of the most important collections in Europe. CRL has acquired several components of the collection, filmed according to 18 broad subject categories. CRL's holdings currently include:

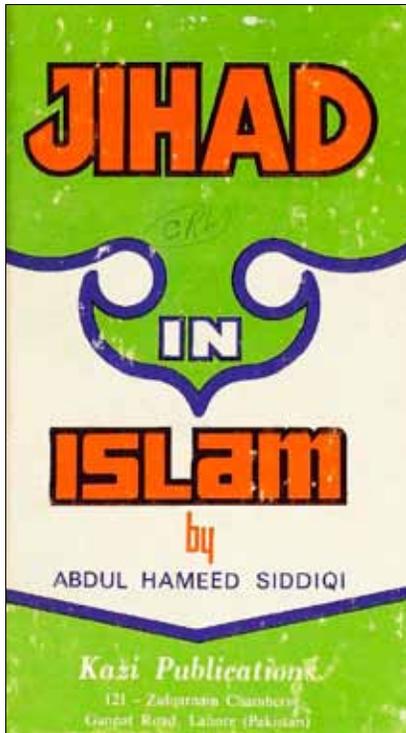
- Ḥadīth: Collections of traditions and statements of Muhammad, along with commentaries and critical works.
- Kalām: theological debates and texts on the Qur'an.
- Islamic mysticism and philosophy: works of Muslim philosophers in all fields of philosophical study. Subsets include:
  - Prayers and sermons
  - Mysticism and pieties
  - Philosophy
  - Ethics and polity.

### Arabic Manuscripts in the Library of the S.O.A.S., University of London

This collection contains more than 400 manuscripts, covering traditional Islamic disciplines such as Tafsir, Hadith, and Fiqh, as well as diverse works on mathematics, astronomy, medicine, falconry, archery, and military equestration. A sizeable portion of the collection relates to Shi'ah literature.

### British Intelligence and Policy on Persia (Iran), c. 1900–49: India Office Political and Secret Files and Confidential Print

This set, filmed from the British Library Oriental and India Office Collections, contains gazetteers and handbooks, military reports, route books, and other printed



The English-language title *Jihād in Islam* by Abdul Hameed Siddiqi (Lahore: Kazi Publications, c. 1979) is available in digital form from the Center for Research Libraries. CRL collections.

correspondence collected by the Government of India Foreign Department and U.K. Foreign Office.

In addition to these major sets, CRL continues to acquire on-demand important doctoral dissertations outside the U.S. and Canada on the subjects of Middle East and Islamic studies. CRL and the British Library recently collaborated to make available online 400 doctoral theses focusing on these subjects. The theses—which represent a wealth of UK postgraduate research into politics, culture, and society in the Middle East, Persian Gulf and North Africa—are freely accessible via the British Library’s [EThOS \(Electronic Theses Online\)](#) service. CRL also purchased numerous doctoral dissertations from Israel in the subject matter of Jewish and Israel Studies.

The Middle East Microform Project (MEMP), CRL’s collaborative program devoted to acquiring and preserving important Middle East resources, continues to build an impressive collection of émigré and opposition newspapers, including publications such as *Agos* (1996–2008), a source for news of Armenian minority in Turkey); *Kayhan* (1984–2010), representing the Iranian community in London as well as other world areas; and *Nur* (2001–08), a Syrian opposition newspaper representing communist and other leftist organizations.

MEMP has also completed the preservation of an important collection of Iranian left-wing materials including pamphlets and other ephemeral material (1941–79) collected by the noted scholar Dr. Cosroe Chaquèri. More recently, MEMP completed the preservation of the Library of Congress Arabic Pamphlet Collection, collected by Dr. George Atiyeh, head of the Near East Section of the Library from 1967 to 1996. The collection includes pamphlets, speeches, and other ephemera collected from across the region. The collection contains more than 4,000 items on 89 microform reels, grouped into 39 major subjects. ❖

*For more details on CRL’s collections, visit the [Topic Guides page on the CRL website](#).*

**FOCUS on Global Resources**, published quarterly, is compiled by CRL's Communications Department. Gwen Ihnat, Don Dyer, Editors. Graphic design services provided by Molly O'Halloran, Inc. Special thanks to James Simon, Gavin Brockett, Robin Dougherty, and Judy Alspach.

ISSN #: 0275-4924

### Center for Research Libraries Staff Contacts (800) 621-6044

President  
Bernard F. Reilly x 334  
breilly@crl.edu

Assistant to the President  
Yvonne Jefferson x 319  
yjefferso@crl.edu

Member Liaison and Outreach  
Services Director  
Mary Wilke x 351  
mwilke@crl.edu

Director of International Resources  
James Simon x 324  
jsimon@crl.edu

Director of Technical Services  
Amy Wood x 327  
awood@crl.edu

Director of Information Systems  
Patricia Xia x 341  
pxia@crl.edu

Head, Access Services  
Kevin Wilks x 314  
kwilks@crl.edu

Head, Stack Management  
Patricia Finney x 328  
pfinney@crl.edu

Digital Program Manager  
Virginia Kerr x 265  
vkerr@crl.edu

Communications Specialist  
Gwen Ihnat x 289  
gihnat@crl.edu

### Global Resources Program Contacts (800) 621-6044

Director  
James Simon x 324  
jsimon@crl.edu

Project Manager—DSAL Project  
Gerald Hall x 318  
ghall@crl.edu

Project Coordinator—  
Global Resources Network  
Judy Alspach x 323  
jalspach@crl.edu

Project Coordinator—Human Rights  
Sarah B. Van Deusen Phillips x 333  
svandeusen@crl.edu

### Contact for Information Regarding: (800) 621-6044

Billing/Invoicing/Deposit Accounts  
Accounting Department x 346

Membership/Communications/  
Public Relations/Marketing  
Don Dyer x 317  
ddyer@crl.edu

CRL Publications Orders (non-*FOCUS*)  
Meeting/Schedules  
Yvonne Jefferson x 319

Loans/Photocopies/Demand  
Purchases  
Access Services Department x 314

Microfilm/Microfiche Sales  
Lorraine Farley x 336  
lfarley@crl.edu

Purchase Proposals  
Mary Wilke x 351

Collection Deposits  
Mary Wilke x 351