Magnetic Needles, Burning Haystacks:
Web-Archiving a Multiethnic Conflict Zone in Russia's North Caucasus

Kit Condill
International & Area Studies Library, University of Illinois at Urbana-Champaign, Urbana, Illinois, U.S.A.
E-mail address: condill@illinois.edu

Abstract:
From the Chechen wars of the 1990s to the Russo-Georgian War of 2008, and continuing through the current spate of political assassinations in Daghestan, Russia’s North Caucasus region has seen more than its share of military, political, ethnic and religious conflict. The past 25 years have also marked a transition from a Soviet-inspired print culture to a vibrant local digital media, which has produced a wealth of competing online content: pro- vs. anti-government; secular vs. religious; tolerant vs. militant; scholarly vs. popular; and perhaps most importantly, content produced within Russian-controlled regions vs. content produced in exile.

The picture is complicated by the extraordinary linguistic diversity of the region, which is home to dozens of distinct languages. My paper will consist, in part, of a survey of online publishing by the ethnic groups of the North Caucasus. Russian, which became the lingua franca of the region in the 20th century, is the main language of publication, but I will also investigate digital media in Chechen, Abkhaz, Ossetian, Avar, Kumyk, and other local languages, which are now struggling to find their place not only in the print and online environments, but also in daily life. Comparisons with earlier periods of upheaval in the North Caucasus, and with conflict zones elsewhere, are possible and will be briefly explored.

Finally, I will present the results of my systematic investigations into the accessibility of online content from the North Caucasus for scholarly research. Websites archived by the University of Illinois will be compared to live websites in the region, and to various commercial products purporting to offer full-text searchability for publications from the North Caucasus. This will provide some quantitative data for the purposes of evaluating the best long-term approaches to collecting and preserving this content for libraries and their users.

Keywords: North Caucasus, web archiving, conflict zones, Chechnya, Dagestan
Introduction

In theory, the rise of the internet and full-text-searchable databases over the past two decades has enabled researchers to find, with just the click of a mouse, what would formerly have taken months or years of painstaking research. In practice, of course, things are much more complicated, especially in the case of a highly multilingual, multiethnic, and conflict-prone region such as the North Caucasus. Searches for relevant materials can be hampered or frustrated completely by choice of search interface, choice of language, choice of Cyrillic vs. Latin scripts, choice of transliteration system, truncation of search terms (or lack thereof), and most of all by the limited amount of North Caucasus news media content that can be searched via any single means (with the partial exception of generic search engines such as Google and Yandex). Ultimately, improved research capabilities for the contemporary North Caucasus are likely to come about only through the efforts of libraries (and/or the vendors that serve them) to unify and preserve the largest possible amount of content, regardless of language, and to provide a robust, effective, and reliable full-text search interface to access that content.

This task is particularly urgent given the historical instability of the North Caucasus, and the not-insignificant likelihood that in the foreseeable future, open hostilities will once again place online, print and manuscript content from the region under grave threat. As the stewards of this content for future generations, librarians and web archivists must strive to prepare in advance for such an eventuality.

Historical context

A brief overview of the current political, military and sociolinguistic situation in the North Caucasus will help to illustrate some of the difficulties associated with such a project. One of the most linguistically and ethnically diverse regions on the planet, the North Caucasus follows the Caucasus Mountains from the Black Sea to the Caspian (see Figs. 1 and 2), and has a long history of resistance to outside rule, which has manifested itself in the last two decades in the brutal Russo-Chechen wars of the 1990s, the Russo-Georgian War of 2008 (which ended with Russia recognizing the independence of the two disputed Georgian regions of Abkhazia and South Ossetia), and present-day terrorist, guerilla, and propaganda activity associated with Islamic State and other fundamentalist movements. During World War II, several North Caucasus peoples were accused of collaborating with the Nazi invaders and deported en masse to Central Asia. Chechnya, for example, was almost completely emptied of Chechens in 1944, and none were allowed to return until the Khrushchev thaw in 1957, only to find that their homes and fields had been occupied by Russians and Ukrainians in the interim, and their print heritage in large part destroyed.1

---

Fig. 1: Southwest Asia. The North Caucasus lies in southern Russian and northern Georgia and Azerbaijan, between the Black Sea and the Caspian.

Fig. 2: Ethnic pluralities in the North Caucasus, by district.

The North Caucasus was fully incorporated into the Russian Empire by the late 19th century, but the more than 30 languages spoken in the region are not related to Russian, and feature some of the most complicated phonological and grammatical features known to linguists – Abkhaz, for example, has 56 consonants and arguably only one vowel, while Chechen has a similar number of consonants and dozens of vowels. In Daghestan, along the
Caspian Sea coast (Fig. 3), Islam arrived in the 7th century, and Arabic served as the language of educated discourse and as a *lingua franca* (one of several) for the remarkable number of ethnolinguistic groups living in its high mountain valleys (see Fig. 4). Repeated waves of invasion and deportation have resulted in substantial North Caucasus diaspora populations in Turkey, Syria, Jordan, Central Asia, and elsewhere over the last two centuries, many of whom have adopted Arabic, Turkish, and other languages as their primary mode of public discourse.

Fig. 3: Location of Dagestan.

Fig. 4: Geographical distribution of Nakh-Daghestani languages within Dagestan.
The Media Environment

Despite concerted (if flawed) efforts to develop newspapers, instructional materials, literary works, and translations of regime-friendly texts in the languages of the North Caucasus during Soviet times, the overall amount of publishing in Chechen, Abkhaz, Circassian (divided by the Soviets into three separate “languages” – Adyge, Cherkess, and Kabardian – in the 1920s), Avar, Karachai-Balkar, Kumyk and other languages has remained small. This continues in the online environment, where a surprising proportion of the available content is in Russian (which, however, is a language much more accessible to Western scholars). This fact is illustrated by one of the few Chechen-language news sources available to Western libraries via a commercial vendor, namely East View’s full-text version of the provincial Chechen newspaper Gums (part of its “Universal Database of the Newspapers of the North Caucasus, Abkhazia and South Ossetia”, a.k.a. UDB-NCA). Although the title of the newspaper is in Chechen (the Russian version of the eponymous city’s name being Gudermes), it turns out that only one of the eight pages in each issue of Gums is in Chechen, and furthermore that that page includes virtually no current affairs or news content, focusing instead on cultural events, poetry by local women poets, and the contents of the latest issues of Chechen literary magazines.

Continuing with East View’s North Caucasus offerings as test cases, we find that less than 25% of the articles in East View’s full-text version of the Vedeno newspaper Kerla dakhar are actually in Chechen. Given that Kerla dakhar has no known website, however (Vedeno has a population of only 3,000 people and is known as one of the most dangerous areas in Chechnya), it is rather remarkable that any content from this local newspaper is available to Western researchers at all. And certainly the Russian-language content does not lack for scholarly interest, as, for example, the struggle of the Chechen and Russian authorities against radical Islam unfolds in real time on its pages. For example, a May 13, 2010 article by R. Agabekov entitled “An emir of the ‘Emirate of the Caucasus’ : not a title, but a (death) sentence” (“Emir ‘Imarata Kavkaz’ : eto ne titul, eto – prigovor”, Kerla dakhar, 2010:19, p. 4) attempts to illustrate the futility of the North Caucasus jihadists’ cause by detailing the deaths of various militants holding the title of “emir” who were killed by Russian security forces in recent months. Later that year, Kerla dakhar announced a three-stage, Chechnya-wide contest for schoolchildren to demonstrate their knowledge of tariqats (traditional Sufi brotherhoods), vird (traditional Sufi prayers), and Wahhabism (or rather, presumably, the spiritual and intellectual bankruptcy thereof), under the guidance of local (government-approved) imams, in order to earn the coveted title of Chechnya-wide “Expert on the foundations of the tariqat” (“Plan provedeniia konkursa ‘Luchshii znatok osnov tarikata’ sredi uchashchikhsia shkol Chechenskoi Respubliki” [entitled by East View as “Konkurs”], Kerla dakhar, 2010:46 [25 Nov.], p. 4).

Other than this small amount of Chechen-language content offered by East View (and, perhaps, some additional content offered via the Integrum database maintained by East View’s competitor, MIPP), all North Caucasus media content known to be commercially available to libraries in Western Europe and North America is in Russian. As noted above, this is not inconsistent with the overall proportion of Russian-language content on the North Caucasian web. Even Vesti respubliki, the official organ of the government of the Republic of Chechnya, is almost entirely in Russian, with the single exception of the stanzas of Chechen-language poetry that are sometimes included in Russian-language articles about literature and culture.

A similar situation exists with regard to communities of North Caucasian origin in the former Ottoman Empire (i.e., present-day Turkey, Syria, Jordan and Iraq) and elsewhere in Europe and Asia. Many of these communities have roots stretching back to the decades-long wars
with the expanding Russian Empire in the mid-19th century, but despite this longevity, it has proven difficult for these communities to maintain any substantial program of publishing in their native languages, most of them adopting Turkish or Arabic as the language of everyday interaction and sociopolitical and cultural engagement (although, especially in Jordan, the place of spoken Chechen as the language of home life is, reportedly, fiercely maintained, although apparently most Jordanian Chechens can neither read nor write in their native language).

So, given this paucity of vernacular-language content vs. the abundance of Russian-language content (even on websites devoted, e.g., to Chechen language, culture, heritage and customs), it is perhaps less surprising to discover that, at the time of this writing, the largest purveyor of news in the vernacular languages of the North Caucasus appears to be the U.S.-government funded Radio Free Europe/Radio Liberty (originally a Cold War propaganda effort overseen by the CIA), with its radio programming and extensive news coverage in Chechen (Marsho Radio, http://www.radiomarsho.com). RFE/RL also provides extensive online reporting in audio, video and prose in Circassian (http://www.khutynygharadio.com/) -- a language divided on the basis of dialect into Adyge, Cherkess and Kabardian in the 1920s -- and Avar (http://www.radioerkenli.com/), the most widely-spoken vernacular language in Dagestan. Marsho Radio’s “Nokhchiin Biblioteka” (i.e., "Chechen Library" -- http://www.radiomarsho.com/a/chechen-online-library/26515068.html) is one of the most promising online repositories for Chechen-language content in the literary and intellectual realm.

RFE/RL’s reporting focuses heavily on corruption, repression, and human rights issues in the North Caucasus, much like one of the region's longest-standing Russian-language news websites: Kavkazskii Uzel, a.k.a. Caucasian Knot (http://www.kavkaz-uzel.eu), which has featured reporting from (usually anonymous) human rights activists from Russia, the North Caucasus, and the South Caucasus for over 15 years. Many more websites based outside the region provide current analysis and less-frequent reporting on events.

There is also an ever-shifting group of jihadist/Salafist websites reporting and commenting on events in the North Caucasus with a pro-Islamic-State slant, usually in Russian. The most long-lasting and stable of these is probably the Finland-based website Kavkaz Center (http://www.kavkazcenter.com/russ/), which has courted controversy with its open support of anti-Russian guerilla movements and terrorism in the North Caucasus (see http://www.rferl.org/content/finnish_kavkaz_center_publisher_vows_to_keep_working_for_chechens/24108353.html). When it comes to openly pro-IS websites and social media forums, of course, there are additional layers of difficulty for librarians and researchers. For example, they tend to be shut down by law enforcement soon after they appear, reappearing later under different titles (much like the repeatedly-censored newspapers of the Bolsheviks and other revolutionaries in the years leading up to the Russian Revolution of 1917).

**Preservation and Searchability**

As will be seen below, in the absence of large-scale or coordinated North Caucasus web archiving programs among public institutions, commercial vendors have an important role to play. While 87% of the websites reviewed for this paper (see Appendix 2) have already been archived at least once via the Internet Archive’s Wayback Machine (https://archive.org/web/), the full text of those sites may be inaccessible due to crawler issues, and in any case cannot be searched without accessing the files via alternative means. In order to investigate the
searchability of full-text media content from the North Caucasus, content produced by the Gudermes newspaper Gums was searched via five different methods:

- East View’s Universal Database of the Newspapers of the North Caucasus, Abkhazia & South Ossetia
- Gums’ own website
- Google search, limited via "site:gums-41.ru"
- Yandex search, limited via "site:gums-41.ru"
- Contents of files archived using SiteSucker

The same 15 search terms were used to test each of the five methods. These search terms (see top row in Table 1) were chosen to illuminate the extent of coverage in both Russian and Chechen for various topics (primarily political and religious, and to a lesser extent, historical and literary), and to provide some sense of the ideological tendency permeating the material (See Appendix I). Other sets of search terms were chosen for other media outlets from other parts of the North Caucasus and beyond, although organizations and events of wide significance (e.g., Beslan and Islamic State) were used throughout.

Table 1: Number of hits for the same search terms using different means of searching the full text of the Chechen regional newspaper Gums

<table>
<thead>
<tr>
<th>Method</th>
<th>Chechen</th>
<th>Russian</th>
<th>History</th>
<th>Literature</th>
<th>Islam</th>
<th>Beslan</th>
<th>Islamic State</th>
<th>Gums</th>
<th>SiteSucker</th>
</tr>
</thead>
<tbody>
<tr>
<td>East View (accessed to 2010)</td>
<td>13</td>
<td>29</td>
<td>2</td>
<td>0</td>
<td>78</td>
<td>0</td>
<td>74</td>
<td>0</td>
<td>115</td>
</tr>
<tr>
<td>Gums-41.ru</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Google (limited to site)</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Yandex (limited to site)</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>39</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>SiteSucker</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Obviously, East View appears to be far and away the most effective means of searching the full text of Gums, but this is primarily due to the fact that East View’s online archive of Gums begins in 2010, while Gums’ own website has no content prior to 2013, and no downloadable full-text content prior to April 2015. It should also be kept in mind that while East View’s database and search interface are extremely useful for the newspaper titles they provide access to, their titles only represent a small subset of the totality of North Caucasus online media
content. As is the case with Integrum, Emerging Markets Information Service (formerly ISI Emerging Markets), and other commercial products, the cost of East View's full-text databases may also be prohibitive or impractical, given the likely level of use they would receive at many institutions.

Equally obviously, Gums' own search capability appears to be seriously flawed at the time of this writing. The major search engines Google and (for the Russian web) Yandex appear to be more effective than the University of Illinois' current means of searching the Gums files it has archived via SiteSucker, but, as major commercial enterprises in their own right, Google and Yandex have a minimal commitment to publicly-accessible archived web content.

**Conclusion**

These relatively minor considerations, however, recede into insignificance when compared to the larger issues of major sites not having been archived yet (by whomever and by whatever means), and the fact that it does not occur to many researchers to make systematic use of this massive and readily-available corpus of primary-source material. Integration of this content into library catalogs or SFX services via appropriate metadata, and making as much full text as possible searchable via commercial or home-grown databases/search interfaces, would seem to be the best path forward, and given the tendency of web content to disappear without warning, it is of some urgency to ensure that as many websites as possible are archived, especially those whose political or social stances and circumstances render them particularly vulnerable. More than elsewhere, unfortunately, the North Caucasus seems to be prone to the catastrophic loss of cultural heritage, and as the possibility of additional conflict and upheaval continues, we, as librarians, do not want to regret not having preserved what we could.
APPENDIX 1: Explanation of search terms for Chechen-related websites

байханов – Ismail Baikhanov (Minister of Education for the Republic of Chechnya, charged with ensuring that education in the republic is pro-regime and anti-Salafist)

беслан – Beslan, North Ossetia (site of the horrific Beslan school hostage crisis of 2004, perpetrated by Chechen separatists)

кадыровцы -- kadyrovtsy (pejorative term for security services beholden to Ramzan Kadyrov, the head of the pro-Russian Chechen government)

кадыровщина -- kadyrovshchina (pejorative term for Kadyrov's regime)


dudiiйн -- Dudiin (Chechen version of Dudaev's last name)

мечеть ахмата кадырова -- Akhmat Kadyrov Mosque (enormous mosque built in Grozny and named after Ramzan Kadyrov's father, claimed to be the largest in Europe)

кадыров ахьмадан цӀарах дина маьждиг -- Chechen name for Akhmat Kadyrov Mosque

сердце чечни -- "Heart of Chechnya" (another name for the Akhmat Kadyrov Mosque)

нохчийчонан дог -- "Heart of Chechnya" in Chechen

ваххабизм -- Wahhabism (Russian term for jihadism/Salafism)

ваххабизм -- Chechen term for Wahhabism/jihadism/Salafism

игил -- IGIL (Russian acronym for the Islamic State in Iraq and the Levant, i.e., IS or ISIL)

исламан пачхаллых -- Chechen term for Islamic State/ISIL

сарақаев -- Ibragim-Bek Sarakaev and/or his son Khamzat Sarakaev, famous Chechen writers and journalists
APPENDIX 2 : Sample list of North-Caucasus-related news, propaganda, and cultural heritage websites

http://www.abkhaziya.org
https://circassianfischt.wordpress.com
http://www.abkhazmoscow.ru/aps
http://apsua.tv
http://инчха.рф
http://www.mashr.org
http://www.shax-dag.ru
http://chechen.org.ua
http://thechechenpress.com
http://interdag.ru
http://www.asiaalwsta.com
http://circassianews.com
http://news.circassianews.com
http://vd.ag
http://www.abkhazeti.info
http://alashara.org
http://apsuaforum.bestforums.org
http://hakikat.info
http://svet-islama05.ru
http://islamdag.ru
http://www.maarulal.ru
http://gergebil.ru
http://azeri.ru
http://forum.bakililar.az
http://haaaman.com
http://j-vaynah.ru
http://kavkazcenter.ru
http://kavkaz-uzel.ru
http://uncukul.ru
http://ботлихра.рф
http://akhvakh.ru
http://minnacrd.ru
http://suleiman-stalskiy.ru
http://samurpress.net
http://magas.ru
http://zamana.info
http://fond-adygi.ru
http://www.adygvoice.ru/wp/
http://apsnvteka.org
http://as-salam.ru
http://khilafa.org
http://khilafa.org/tag/fisyria-com/
http://vesti95.ru/
http://gums-41.ru/
http://zama.fm/

http://www.nohchalla.com/ -- Chechen cultural heritage/cultural pride -- some Chechen-language content, significant Chechen language-learning aids (audio and written) -- many articles on the deportation -- huge section on the "microtoponymy" of Chechnya

http://nana-journal.ru/
http://www.vainahkrgrk.kz/ -- website for Chechen cultural centers in diaspora in Kazakhstan (i.e., for victims of deportation that did not choose to return to Chechnya after 1957) -- all content appears to be in Russian

https://hutba.org/ -- where, within Russia, the website is based is (intentionally?) unclear -- provides assistance in learning Arabic and basic Islamic traditions and principles -- "the first online madrasa"

http://www.chechensinsyria.com

http://www.chechnyaadvocacy.org

http://marshodergisi.blogspot.com

http://marshodergisi.blogspot.com/p/dergi-arsivi.html

http://chechen.org/ -- although full of empty pages and difficult to navigate, this website provides access to the full text of the journal Prometheus, which focuses on historical and cultural studies