

Data-Driven Citizen Journalism

Case Study: The Afghan War Diary Released by WikiLeaks

By Camilo Cristancho, Catalina Iorga and Matteo Cernison

Digital Methods Summer School 2010

Foundations for Online Research with Digital Methods

Research Question

Is there an alternative account of the Afghan War Diary 2004 - 2010 documents released by WikiLeaks? In other words, is the data hosted by WikiLeaks used in different ways other than those employed by the mainstream media represented by WikiLeaks' official partners?

Is data-driven citizen journalism present on the Web? If so, how does it manifest? What are the content-syndication practices of alternative voices?

Retrieving Inlinks to Afghan War Diary Individual Documents

1. Observe which is the common root of all document URLs, namely 'http://wardiary.wikileaks.org/afg/event'.
2. Query Google by using the **Google Scraper** to obtain the first 1000 results which contain this common root as a textual component.
3. Submit the 95 obtained webpages (alternatively considered as the top 100) to the **Link Ripper** in order to later get all outlinks to specific **Afghan War Diary 2004 - 2010** document pages.
4. Insert the **Link Ripper** output in the **Harvester** in order to alphabetize the obtained URLs and remove textual descriptions.
5. Manually clean the output by again searching for the 'http://wardiary.wikileaks.org/afg/event' in an Excel file and produce a separate list of **Afghan War Diary 2004 - 2010** document URLs.
6. Analyze the list containing 179 non-unique results, select all document pages that receive at least two links and create a file with the 'most mentioned' warlogs, 17 to be exact.

Which Alternative Voices Cite Which Popular Documents

1. Idaho Soldier Captured in Afghanistan

wardiary.wikileaks.org/afg/event/2009/12/AFG20091208n2517.html

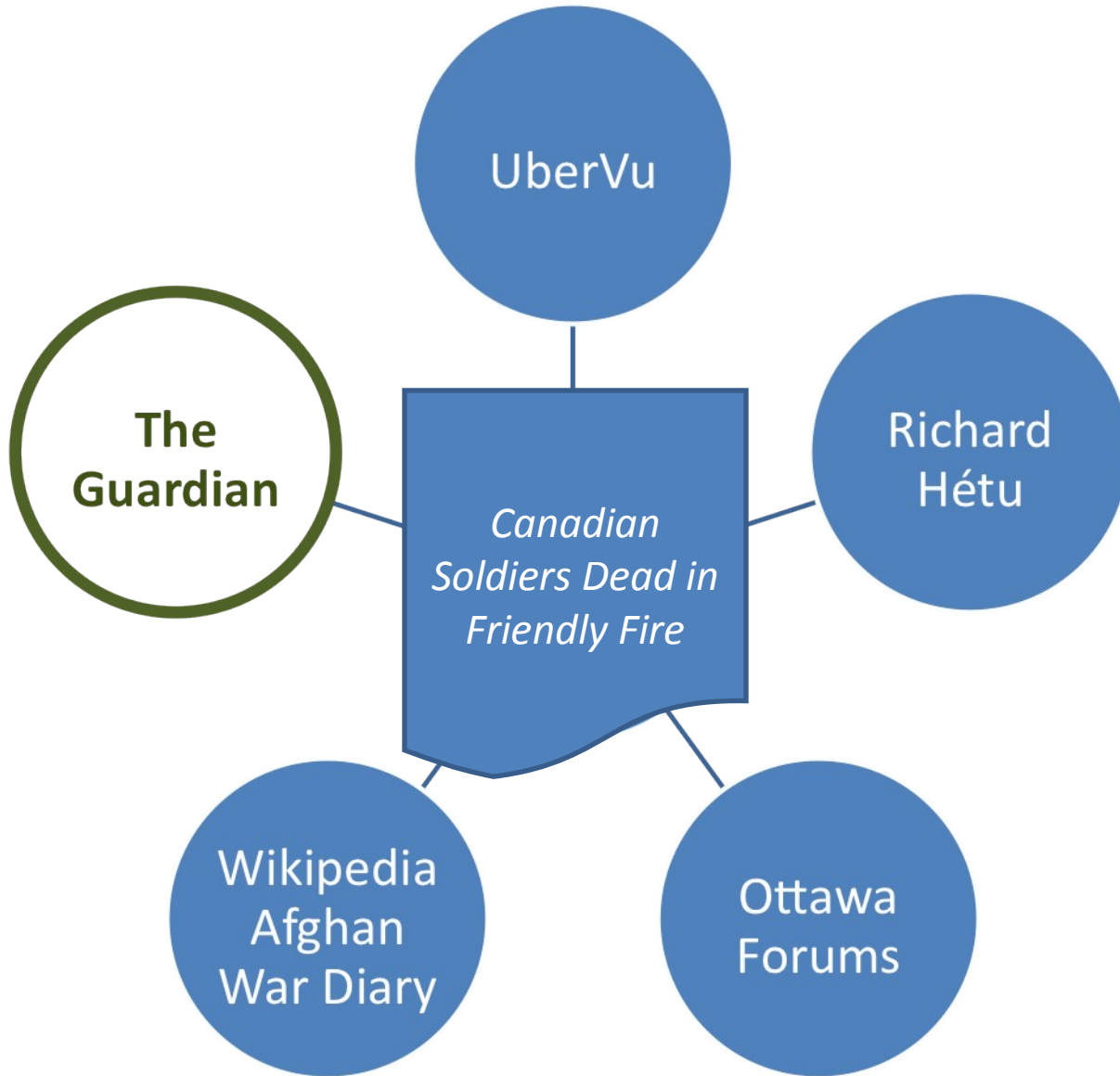
- Cited by no alternative outlets

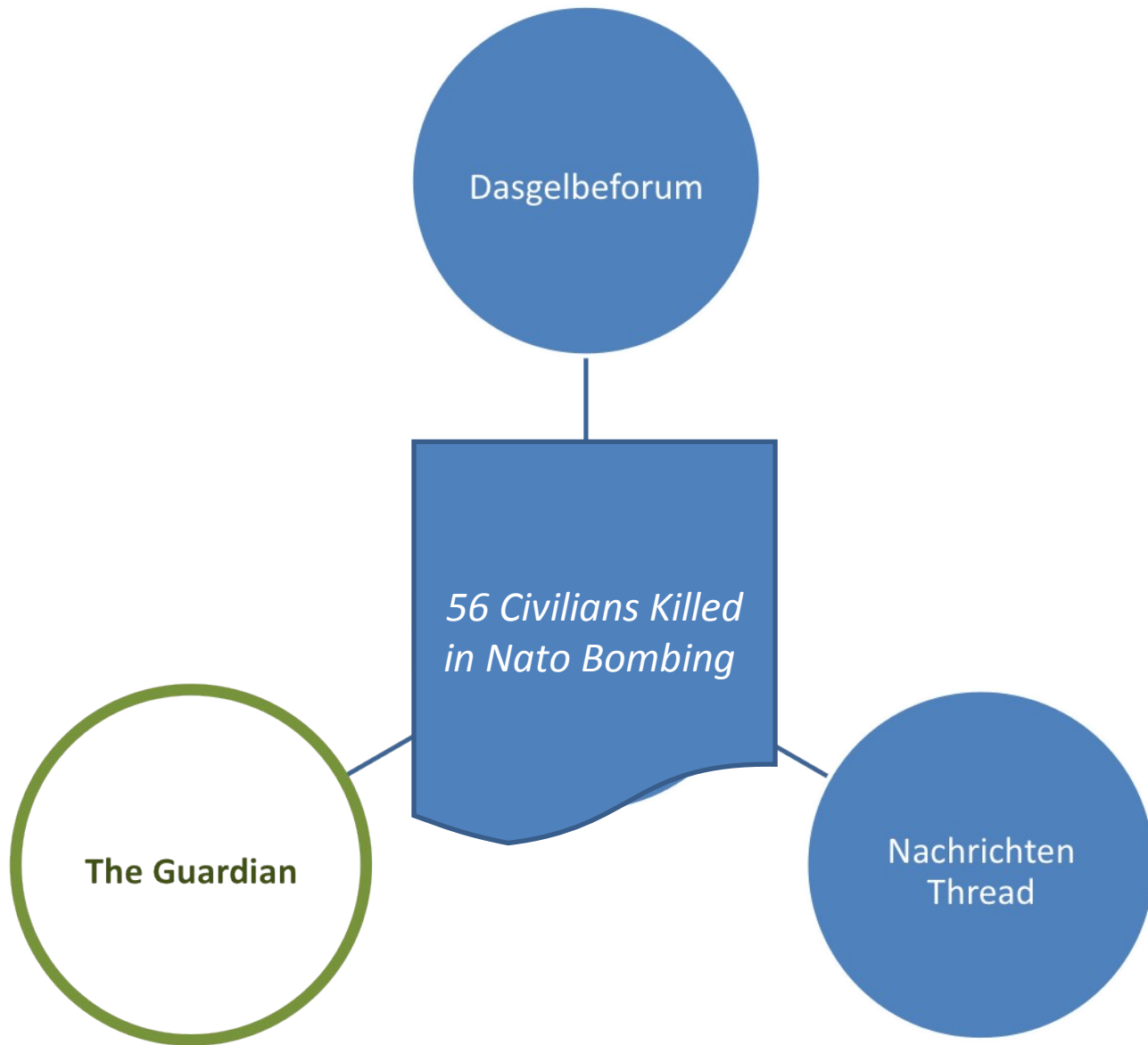
2. Four Canadians Killed in Friendly Fire

wardiary.wikileaks.org/afg/event/2006/09/AFG20060903n347.html

- Mentioned on Ottawa Forums (July 26 2010) and presented as a 'hot topic'
- Present in the Comment Section of Richard Hétu's Blog (July 27 2010) and framed as US cover-up attempt





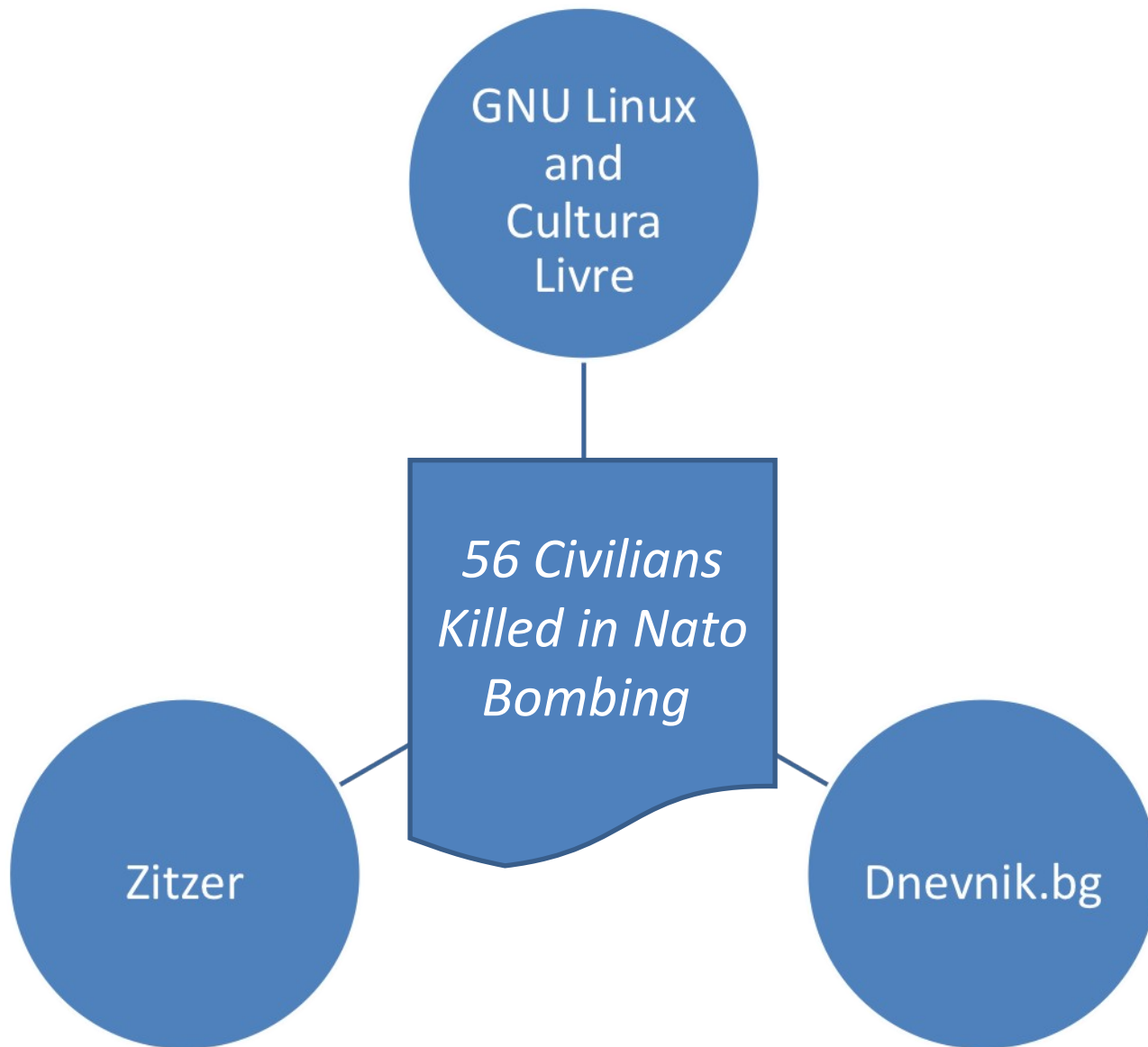


Dasgelbeforum

*56 Civilians Killed
in Nato Bombing*

The Guardian

Nachrichten
Thread



GNU Linux
and
Cultura
Livre

*56 Civilians
Killed in Nato
Bombing*

Zitzer

Dnevnik.bg

Verkkouutis
e.com

*4 Wounded
Finnish
Soldiers*

*Killed Finnish
Soldier*

Verkkouutis
e.fi

Websites Linking to wardiary.wikileaks.org

Digital Methods Initiative

13 August

10

Map generated by tools.digitalmethods.net

dialogic.blogspot.com (270) freedetainees.org (149)

adoptresistance.blogspot.com (26) huffingtonpost.com (14) labvirus.wordpress.com (12) salon.com (12) guardian.co.uk (11) bbc.co.uk (11) mediavigil.blogspot.com (9) seeker401.wordpress.com (6)

trinityatierra.wordpress.com (6) cryptome.org (4) docuticker.com (4) fayerwayer.com (4) cfr.org (4) antiwar.com (4) neufneuf.blogspot.com (4) newyorker.com (4) committeetoprotectbloggers.org (4)

english.aljazeera.net (3)

Websites Linking to wardiary.wikileaks.org/afg/sort/type

Digital Methods Initiative

13 August

10

Map generated by tools.digitalmethods.net

forum.opiophile.org (4)
tangibleinfo.blogspot.com (3) tumblr.com (2)

just-another-inside-job.blogspot.com (2) offnews.info (2) wired.com (2) aeroweb-fr.net (1) evangelisch.de (1)

trackpads.com (1) abovetopsecret.com (1) rss-verzeichnis.de (1) vott.ru (1) mcnorman.wordpress.com (1) islandbreath.blogspot.com (1) politics.gather.com (1) smledbetter.com (1) blazingcatfur.blogspot.com (1)

tidal.heroku.com (1) webcommunity.ilvolo.it (1)

The Afghan War Diary in the Twittersphere

Digital Methods Initiative

13 August

10

Map generated by tools.digitalmethods.net

#wikileaks (3923)

wardiary.wikileaks.org (2024) #warlogs (346) AfghanWarDiary (163) #wardiary (9)