# Dan Cohen's Digital Humanities Blog » Blog Archive » Mapping What Americans Did On September 11

I gave a talk a couple of days ago at the annual meeting of the Society for American Archivists[1] (to a great audience—many thanks to those who were there and asked such terrific questions) in which I showed how researchers in the future will be able to intelligently search, data mine, and map digital collections. As an example, I presented some preliminary work I've done on our September 11 Digital Archive[2] combining text analysis with geocoding to produce overlays on Google Earth[3] that show what people were thinking or doing on 9/11 in different parts of the United States. I promised a follow-up article in this space for those who wanted to learn how I was able to do this. The method provides an overarching view of patterns in a large collection (in the case of the September 11 Digital Archive, tens of thousands of stories), which can then be prospected further to answer research questions. Let's start with the end product: two maps (a wide view and a detail) of those who were watching CNN on 9/11 (based on a text analysis of our stories database, and colored blue) and those who prayed on 9/11 (colored red).
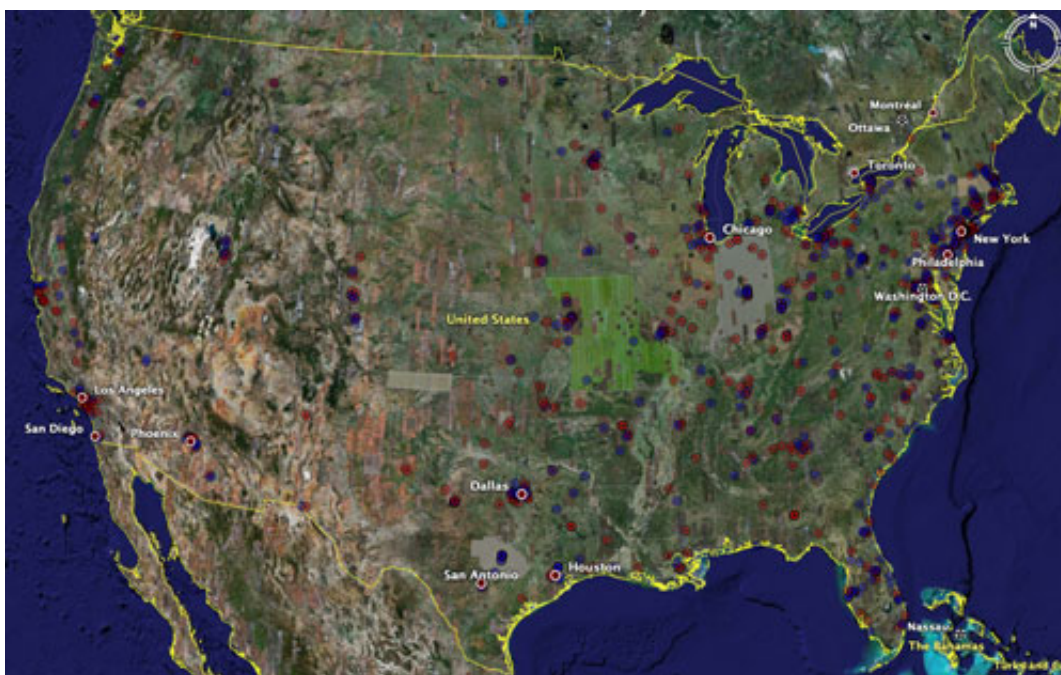
Google Earth map of the United States showing stories with CNN viewing (blue) and stories with prayer (red) [view full-size version for better detail[4]]
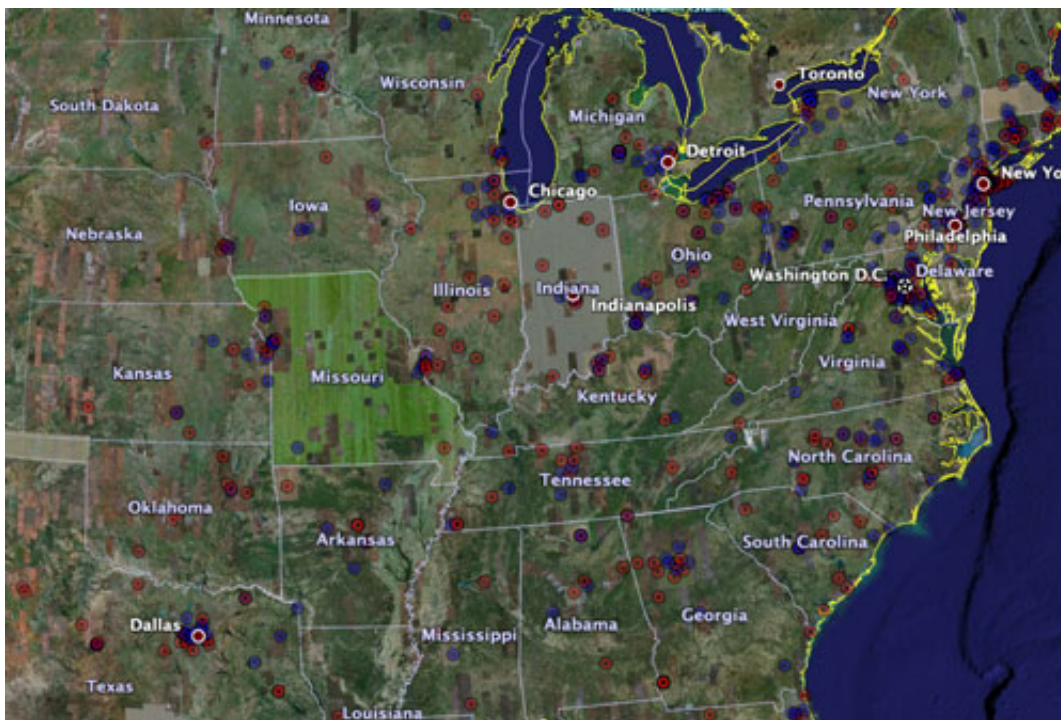
Detail of the Eastern United States [view full-size version for better detail[5]]

By panning and zooming, you can see some interesting patterns. Some of these patterns may be obvious to us, but a future researcher with little knowledge of our present could find out easily (without reading thousands of stories) that prayer was more common in rural areas of the U.S. in our time, and that there was especially a dichotomy between the very religious suburbs (or really, the exurbs) of cities like Dallas and the mostly urban CNN-watchers. (I'll present more surprising data in this space as we approach the fifth anniversary of 9/11.)

OK, here's how to replicate this. First, a caveat. Since I have direct access to the September 11 Digital Archive database, as well as the ability to run server-to-server data exchanges with Google and Yahoo (through their

API programs), I was able to put together a method that may not be possible for some of you without some programming skills and direct access to similar databases. For those in this blog's audience who do have that capacity, here's the quick, geeky version: using regular expressions, form an SQL query into the database you are researching to find matching documents; select geographical information (either from the metadata, or, if you are dealing with raw documents, pull identifying data from the main text by matching, say, 5-digit numbers for zip codes); put these matches into an array, and then iterate through the array to send each location to either Yahoo's[6] or Google's[7] geocoding service via their maps API; take the latitude and longitude from the result set from Yahoo or Google and add these to your array; iterate again through the array to create a KML[8] (Keynote Markup Language) file by wrapping each field with the appropriate KML tag[9].

For everyone else, here's the simplest method I could find for reproducing the maps I created. We're going to use a web-based front end for Yahoo's geocoding API, Phillip Holmstrand's very good free service[10], and then modify the results a bit to make them a little more appropriate for scholarly research.

First of all, you need to put together a spreadsheet in Excel (or Access or any other spreadsheet program; you can also just create a basic text document with columns and tabs between fields so it looks like a spreadsheet). Hopefully you will not be doing this manually; if you can get a tab-delimited text export from the collection you wish to research, that would be ideal. One or more columns should identify the location of the matching document. Make separate columns for street address, city, state/province, and zip codes (if you only have one or a few of these, that's totally fine). If you have a distinct URL for each document (e.g., a letter or photograph), put that in another column; same for other information such as a caption or description and the title of the document (again, if any). You don't need these non-location columns; the only reason to include them is if you wish to click on a dot on Google Earth and bring up the corresponding document in your web browser

(for closer reading or viewing).

Be sure to title each column, i.e., use text in the topmost cell with specific titles for the columns, with no spaces. I recommend "street_address," "city," "state," zip_code," "title," "description," and "url" (again, you may only have one or more of these; for the CNN example I used only the zip codes). Once you're done with the spreadsheet, save it as a tab-delimited text file by using that option in Excel (or Access or whatever) under the menu item "Save as…"

Now open that new file in a text editor like Notepad on the PC or Textedit on the Mac (or BBEdit or anything else other than a word processor, since Word, e.g., will reformat the text). Make sure that it still looks roughly like a spreadsheet, with the title of the columns at the top and each column separated by some space. Use "Select all" from the "Edit" menu and then "Copy."

Now open your web browser and go to Phillip Holmstrand's geocoding website[11] and go through the steps. "Step #1″ should have "tab delimited" selected. Paste your columned text into the big box in "Step #2″ (you will need to highlight the example text that's already there and delete it before pasting so that you don't mingle your data with the example). Click "Validate Source" in "Step #3." If you've done everything right thus far, you will get a green message saying "validated."

In "Step #4″ you will need to match up the titles of your columns with the fields that Yahoo accepts, such as address, zip code, and URL. Phillip's site is very smart and so will try to do this automatically for you, but you may need to be sure that it has done the matching correctly (if you use the column titles I suggest, it should work perfectly). Remember, you don't need to select each one of these parameters if you don't have a column for every one. Just leave them blank.

Click "Run Geocoder" in "Step #5″ and watch as the latitudes and longitudes appear in the box in "Step #6." Wait until the process is totally done. Phillip's site will then map the first 100 points on a built-in

Yahoo map, but we are going to take our data with us and modify it a bit. Select "Download to Google Earth (KML) File" at the bottom of "Step #6." Remember where you save the file. The default name for that file will be "BatchGeocode.kml". Feel free to change the name, but be sure to keep ".kml" at the end.

While Phillip's site takes care of a lot of steps for you, if you try right away to open the KML file in Google Earth you will notice that all of the points are blazing white. This is fine for some uses (show me where the closest Starbucks is right now!), but scholarly research requires the ability to compare different KML files (e.g., between CNN viewers and those who prayed). So we need to implement different colors for distinct datasets.

Open your KML file in a text editor like Notepad or Textedit. Don't worry if you don't know XML or HTML (if you do know these languages, you will feel a bit more comfortable). Right near the top of the document, there will be a section that looks like this:

```
<Style id="A"><IconStyle><scale>0.8</scale><Icon><href>root://icons/palette-4.png</href><x>30</x><w>32</w><h>32</h></Icon></IconStyle><LabelStyle><scale>0</scale></LabelStyle></Style>
```

To color the dots that this file produces on Google Earth, we need to add a set of "color tags" between <IconStyle> and <scale>. Using your text editor, insert "<color></color>" at that point. Now you should have a section that looks like this:

```
<Style id="A"><IconStyle><color></color><scale>0.8</scale><Icon><href>root://icons/palette-4.png</href><x>30</x><w>32</w><h>32</h></Icon></IconStyle><LabelStyle><scale>0</scale></LabelStyle></Style>
```

We're almost done, but unfortunately things get a little more technical. Google uses what's called an ABRG value for defining colors in Google Earth files. ABRG stands for "alpha, blue, green, red." In other words, you will have to tell the program how much blue, green, and red you want in the color, plus the alpha value, which determines how opaque or transparent the dot is. Alas, each of these four parts must be expressed in a two-digit hexidecimal format ranging from "00″ (no amount) to "ff" (full amount). Combining each of these two-digit values gives you the necessary full string of eight characters. (I know, I know—why not just <color>red</color>? Don't ask.) Anyhow, a fully opaque red dot would be <color>ff00ff00</color>, since that value has full ("ff") opacity and full ("ff") red value (opacity being the first and second places of the eight characters and red being the fifth and sixth places of the eight characters). Welcome to the joyous world of ABRG.

Let me save you some time. I like to use 50% opacity so I can see through dots. That helps give a sense of mass when dots are close to or on top of each other, as is often the case in cities. (You can also vary the size of the dots, but let's wait for another day on that one.) So: semi-transparent red is "7f00ff00″; semi-transparent blue is "7fff0000″; semi-transparent green is "7f0000ff"; semi-transparent yellow is "7f00ffff". (No, green and red don't make yellow, but they do in this case. Don't ask.) So for blue dots that you can see through, as in the CNN example, the final code should have "7fff0000″ inserted between <color> and </color>, resulting in:

<Style id="A"><IconStyle><color>7fff0000</color><scale>0.8</scale><Icon><href> root://icons/palette-4.png</href><x>30</x><w>32</w><h>32</h></Icon> </IconStyle><LabelStyle><scale>0</scale></LabelStyle></Style>

When you've inserted your color choice, save the KML document in your text editor and run the Google Earth[12] application. From within that application, choose "Open…" from the "File" menu and select the KML

file you just edited. Google Earth will load the data and you will see colored dots on your map. To compare two datasets, as I did with prayer and CNN viewership, simply open more than one KML file. You can toggle each set of dots on and off by clicking the checkboxes next to their filenames in the middle section of the panel on the left. Zoom and pan, add other datasets (such as population statistics), add a third or fourth KML file. Forget about all the tech stuff and begin your research.

[For those who just want to try out using a KML file for research in Google Earth, here are a few from the September 11 Digital Archive. Right-click (or control-click on a Mac) to save the files to your computer, then open them within Google Earth, which you can download from here[13]. These are files mapping the locations of: those who watched CNN[14]; those who watched Fox News[15] (far fewer than CNN since Fox News was just getting off the ground, but already showing a much more rural audience compared to CNN); and those who prayed[16] on 9/11.]

This entry was posted on Tuesday, August 8th, 2006 at 4:32 pm and is filed under Archives[17], Google[18], Maps[19], Research[20], Tutorials[21], Visualization[22]. You can follow any responses to this entry through the RSS 2.0[23] feed. You can leave a response[24], or trackback[25] from your own site.

References

1. ^ Society for American Archivists (www.archivists.org)
2. ^ September 11 Digital Archive (911digitalarchive.org)
3. ^ Google Earth (earth.google.com)
4. ^ view full-size version for better detail (www.dancohen.org)
5. ^ view full-size version for better detail (www.dancohen.org)
6. ^ Yahoo's (developer.yahoo.com)
7. ^ Google's (www.google.com)
8. ^ KML (earth.google.com)
9. ^ KML tag (earth.google.com)
10. ^ Phillip Holmstrand's very good free service (www.batchgeocode.com)

11. ^ Phillip Holmstrand's geocoding website (www.batchgeocode.com)
12. ^ Google Earth (earth.google.com)
13. ^ download from here (earth.google.com)
14. ^ those who watched CNN (www.dancohen.org)
15. ^ those who watched Fox News (www.dancohen.org)
16. ^ those who prayed (www.dancohen.org)
17. ^ View all posts in Archives (www.dancohen.org)
18. ^ View all posts in Google (www.dancohen.org)
19. ^ View all posts in Maps (www.dancohen.org)
20. ^ View all posts in Research (www.dancohen.org)
21. ^ View all posts in Tutorials (www.dancohen.org)
22. ^ View all posts in Visualization (www.dancohen.org)
23. ^ RSS 2.0 (www.dancohen.org)
24. ^ leave a response (www.dancohen.org)
25. ^ trackback (www.dancohen.org)

Excerpted from *Dan Cohen's Digital Humanities Blog » Blog Archive » Mapping What Americans Did*

on September 11

http://www.dancohen.org/2006/08/08/mapping-what-americans-did-on-september-11/