

## MAPPING THE NEWS WITH OPENBLOCK – PRESENTING LOCAL NEWS GEOGRAPHICALLY

Stephen Nowers, University of Alaska Anchorage, [stephen.nowers@gmail.com](mailto:stephen.nowers@gmail.com)  
Yi-Miin Minnie Yen, University of Alaska Anchorage, [afmyy@uaa.alaska.edu](mailto:afmyy@uaa.alaska.edu)  
Yoshito Kanamori, University of Alaska Anchorage, [afyk@uaa.alaska.edu](mailto:afyk@uaa.alaska.edu)

### ABSTRACT

A vast amount of government data is available digitally but not in a form useful to the public. Online news entities, in particular Everyblock.com, are beginning to present this data online but not in an up-to-the-minute form and not for mid-size communities like Anchorage. Additionally, there is no doubt that traditional news organizations are concerned about the loss of viewers and circulation. This project attempts to solve these problems and explore the potentials in the emerging field of computational journalism. The Anchorage OpenBlock project has been trying to follow EveryBlock's footsteps and rationale with a "geographic filter" that provides a "news feed" for users' neighborhoods.

**Keywords:** Computational Journalism, OpenBlock, EveryBlock, block news, hyper-local service

### INTRODUCTION

In his talk at Hong Kong University Journalism and Media Center in March, 2011, Dr. Nick Diakopoulos defined Computational Journalism as "using computing to facilitate, enable and reinvigorate the practices and processes of journalism, including collecting, organizing, making sense of, communicating and disseminating news information, while upholding the values of journalism such as fairness and accuracy" [1],[2]. He argues that computation (and the related networking, information technology, etc.) is now intertwined with all aspects of journalism, including newsgathering, investigative journalism, verification/fact finding, authoring/printing/publication/broadcasting of news, sharing and distribution of news information, editing and commenting on news, etc [3].

Computation has not only impacted these aspects, but has also thoroughly disrupted journalism's traditional distribution models. Journalists now have new ways of providing the information citizens need for civic engagement. In other words, the major objective of Computational Journalism is to study the overlapping interests between computation and journalism and to define how both of these elements can help with information gathering and dissemination for and by

citizens to achieve an engaged and more actively participating citizenry.

As Diakopoulos mentioned in his talk, "information comes at us faster and faster and the news media have more and more data to deal with. Social media is pumping out terabytes of this every day. We need computers to help us deal with that scale." We also see that a vast amount of government data is available digitally but not in a form useful to the public. Although the general public still believes that newspapers play a critical role in the democratic process by framing the conversation and encouraging informed debate and decision-making, there is no doubt that traditional news organizations are concerned about the loss of viewers and circulation. And this issue is even more acute when it comes to local news. For example, the largest newspaper in Alaska, the Anchorage Daily News, has lost at least 115 positions company-wide through layoffs and attrition since June 2008 [4],[5],[6]. Newspaper companies have struggled in the age of the Internet, but this disruptive medium also offers opportunities for a new kind of journalism – one that can play the same critical role in the society and do it more efficiently.

With this in mind we began a computational journalism project to collect and present localized public data to users in Anchorage, Alaska. Our effort follows in the footsteps of the 2007 Knight News Challenge winning EveryBlock project.

Our goal is to adapt EveryBlocks's code and experiment with new ways for users to find local news and public information. This project will be implemented in two phases: the first phase started with the installation of a pilot version -- including mapping and updated news feeds -- customized for our community in Anchorage. The objectives we have set up for this project's news feeds should be:

- a. Automated. We need to use data primarily through automated feeds. Due to limited time available, data collection and presentation need to be accomplished through scripting.
- b. Trustworthy. The source of the data must be included in the project site design.

- c. Timely. Delayed information will severely limit the usefulness of the project.
- d. Scalable. The capacity to add data feeds as they are identified needs to be built into the design of the project.

The second phase of the project will focus on the sustainability of this new form of online news site. This phase will experiment and explore a workable business model, with a sustainable, and even profitable, sit our long-term goal. It will proceed after we finish the first phase to get the site up and running. This paper primarily discusses the work accomplished in the first phase. It starts with an introduction section and is followed by the sections on history and background, technology and implementation, conclusion and future study.

## HISTORY and BACKGROUND

Federal, state, and local government agencies publish vast amount of public data online but much of it is in a form that is difficult for consumers to use. News organizations are starting to explore this space, but they either lack the technological staff in their organization to process the data or else tend to build interactive, Flash-based graphics for a specific story, such as the New York Times's Broad Unemployment Across the U.S. [8] and Clean Water Act Violations: The Enforcement Record [9]. These data packages are usually difficult and time consuming for news organizations to create. It usually requires three personnel assets to produce this kind of data graphic: expert reporters, data integration expertise and Flash graphic expertise. The New York Times might be the last newspaper in North America that has all three, according to Mindy McAdams at the University of Florida [10].

Some non-traditional journalism sites like the political site FiveThirtyEight [11] and news mapping site EveryBlock [12] have entered this data-driven reporting space but they offer the data of interest to only a small segment of users. EveryBlock.com (figure 1) is a new experiment in online journalism, collecting public data for 16 American major cities and presenting the information for online by neighborhood and zip code. It began with a two-year, \$1.1 million grant funded by the Knight Foundation [13] and was acquired by MSNBC.com in August 2009 [14].

The goal of EveryBlock is to collect all of the news and civic goings-on that have happened recently in the user's block, and make it simple for users to keep track of news in particular areas. They seek to be a

geographic filter – a “news feed” for customer's neighborhood or block[15]. A user need only enter an address, neighborhood or ZIP code in the cities being served, and the site will produce recent public records, news articles and other Web content that's geographically relevant to the location.

At EveryBlock, they treat freshly updated public records as block-level news. It's their goal to acquire ongoing feeds of government data and spread the gospel of open data. The two major objectives are to get more datasets for EveryBlock so it can be a better Web site and to convince governments to share that data with everyone, not just the site itself.

In June 2010, the John S. and James L. Knight Foundation announced a plan to expand the EveryBlock project to make it easier for news organizations to use the EveryBlock code [16]. The new software, which is also an open source, is called OpenBlock. The Knight Foundation has awarded nearly \$460,000 in this second round of funding. OpenPlans, a New York-based non-profit organization received a \$235,000 grant to extend the EveryBlock.com code while the Columbia Daily Tribune and the Boston Globe were awarded with grants to install and test the software. OpenPlans began releasing updates to the Everyblock code in September, 2010 and the chief developers are active on the EveryBlock discussion group (eb\_code) on Google Groups [17]. The OpenBlock is a project currently under active development, with a public release planned for Spring 2011.

## TECHNOLOGY and IMPLEMENTATION

### A. Technologies

The project requirements in large part dictated the core technology choices. Our EveryBlock Anchorage project is based on the Django Web Framework [18] by following the technologies chosen by EveryBlock (now OpenBlock). Django is an open-source framework developed for the online operation of a newspaper which can be used to create Web applications on tight deadlines. “[Django is] is programming tool that lets you build web sites and uses the Python programming language.” [19]. We chose the Linux operating system Ubuntu (desktop V9.10) for simplicity. This project also requires PostGIS, an open-source spatial database extender for PostgreSQL.

This project adopts a parallel desktop virtual machines approach for its development to insure a clean development environment and allow snapshots to be taken as progress is made or

whenever the dependencies and libraries are added and updated.

## B. Implementation Methodology

### 1. Install instance for Anchorage

The project started to use the EveryBlock open-source code with a clean installation of Ubuntu 9.04 and all updates. In-depth research was needed in addition to the original EveryBlock set-up instructions. With support from help documents for GeoDjango [20], a geospatial extension of Django, we finally got the original framework done. During this process, the eb\_code group on Google Groups also provided critical assistance, including installation instructions for the EveryBlock code on Ubuntu 9.10, as well as the steps to load demonstration data for the city of Boston.

The EveryBlock community expanded dramatically in the summer of 2010 with the Knight Foundation's funding of the OpenBlock project. The Open Block team forked an existing github EveryBlock repository [21] and released an alpha version of their code in September [22]. The OpenBlock repository includes a map component, a Python-based bootstrap file in order to load Boston-area data into the application, as well as the ability to use Django's GUI administration interface.

### 2. Loading local data

Data must be added once the application is properly installed. The U.S. Census Bureau provides shape files containing streets and intersections for cities across the country [23]. These are the suggested 'blocks' in EveryBlock. Our Anchorage project used these Census Bureau's shape files for our installation.

EveryBlock and OpenBlock use the same Python script to import block data into the database. This file, `import_blocks.py`, collects primary roads, secondary roads, city streets, alleys, service roads and private roads from 'edges\_shp' and 'featnames\_dbf' from the U.S. Census Bureau database (the Anchorage files are 'tl\_2009\_02020\_edges.shp' and 'tl\_2009\_02020\_featnames.dbf').

The application also requires news items to display. This project used the lifestyles (entertainment) RSS feed from the Anchorage Daily News [24]

To create the EveryBlock Anchorage site requires two separate steps: installing a working instance of the application and loading local data into the application.

and the Anchorage RSS feed from SeeClickFix.com [25]. The successful installation of EveryBlock: Anchorage adapted the Boston-specific command line instructions provided on the eb\_code group on Google Groups to Anchorage. The initial installation of OpenBlock: Anchorage adapted the Boston-specific files and scripts packaged with the OpenBlock: Boston demonstration package. The current installation of OpenBlock: Anchorage uses a clean installation of OpenBlock (v0.1) with custom files and scripts.

### 3. Updating OpenBlock's extension of the Everyblock source code

The installation of EveryBlock turned out to be a difficult undertaking, primarily because of the way it handles mapping. EveryBlock uses a separate module, `eb_geo`, to map news items. `eb_geo` renders map tiles locally using the free C++-based toolkit Mapnik, which requires a fairly complicated set of interconnected libraries and plug-in dependencies. The `eb_geo` manual itself warns "the installation of Mapnik is non-trivial" [26].

Once `eb_geo` is installed, developers must create their own map tiles in Mapnik-style XML. The templates included with `eb_geo` do not include any maps at all at this time. Though an EveryBlock – Anchorage instance was successfully created during the course of this project, it did not include the mapping function. The Knight Foundation also addressed the installation difficulties by funding the OpenBlock project to ease the creation of EveryBlock instances [27]. OpenBlock resolves mapping difficulties by using a free base layer based on Open Street Map and hosted by OpenGeo. This bypasses the need for both the `eb_geo` module and Mapnik, and thus eliminating many of the problematic library dependencies for EveryBlock. This base layer is not hosted locally and the application must be able to reach Open Street Map to be able to serve the maps. This part of our project is a work-in-progress, which we will develop using OpenBlock's Open Street Map technique for mapping.

The OpenBlock v0.1 code also simplified installation with bootstrap files for the software and a demo site (OpenBlock Demo: Boston).

Bootstrap.py sets up a virtual environment and installed the Openblock software and its Python requirements while bootstrap\_demo.sh loads the geographies for the city of Boston, schema definitions and some news. The scripts called by bootstrap\_demo.sh, including import\_boston\_hoods.sh, add\_boston\_news\_schemas.sh and import\_boston\_news.sh, are easy to customize for a different municipality.

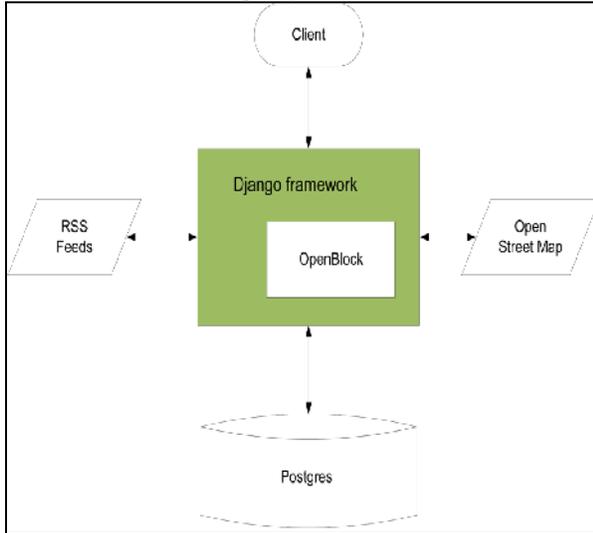


Figure 1: OpenBlock interface diagram

### C. OpenBlock Anchorage Results

Originally, this project had installed the EveryBlock open-source code successfully and loaded with the Anchorage TIGER/Line data and a RSS feed from the Anchorage Daily News in April, 2010 (see Figure 2). Adopting the OpenBlock code fork greatly simplifies the installation process and adds additional features into our implementation. Once the components required by OpenBlock are installed and configured on the selected operating system, the sample demo Boston site of OpenBlock v0.1 is operational with three shell commands [28]. The local demo for the city of Anchorage is similar to the OpenBlock Development Demo site and is fully functional. The updated OpenBlock Anchorage site relies on the demo site's bootstrap\_demo.sh scripts with

Django, the web framework used by EveryBlock, includes a powerful administration console that was left out of the EveryBlock source code. The OpenBlock project adds this utility to the application. See OpenBlock interface diagram (Figure 1).

calls to Anchorage-specific scripts in order to load data (see Figure 3).

Though the application itself is much easier to install, it still requires a local data feed. Data from the Anchorage Daily News features section RSS feed and the SeeClickFix GeorSS feed were loaded into the application but only those from SeeClickFix contain enough information, such as latitude and longitude points, to allow mapping at this time. The administration site was also implemented and provided a simple GUI to modify the elements of its own database (see Figure 4).

### CONCLUSION AND FUTURE STUDY

Although the OpenBlock project is still an alpha release, the code fork is already much easier to use and developers interested in expanding to other cities can begin to turn their attention to collecting the local data necessary to make spin-off sites useful for their municipalities. Our Anchorage OpenBlock project created a site which demonstrated how EveryBlock code, by using the OpenBlock fork, can automatically import data from a major media source into a local application in a timely manner. The application's scope will expand as more data feeds and more available functionalities are added to the site.

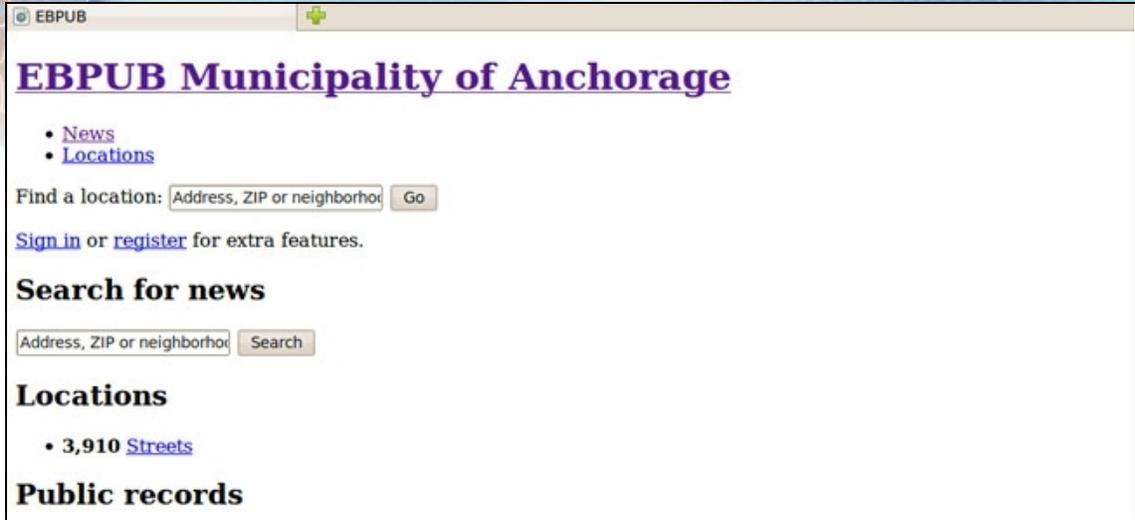


Figure 2: EveryBlock Anchorage homepage

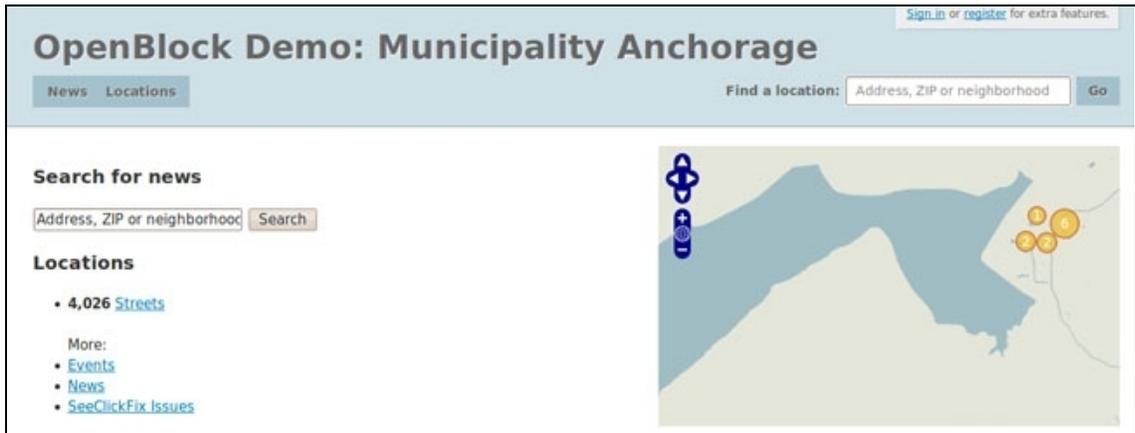


Figure 3: OpenBlock Demo (Anchorage) homepage

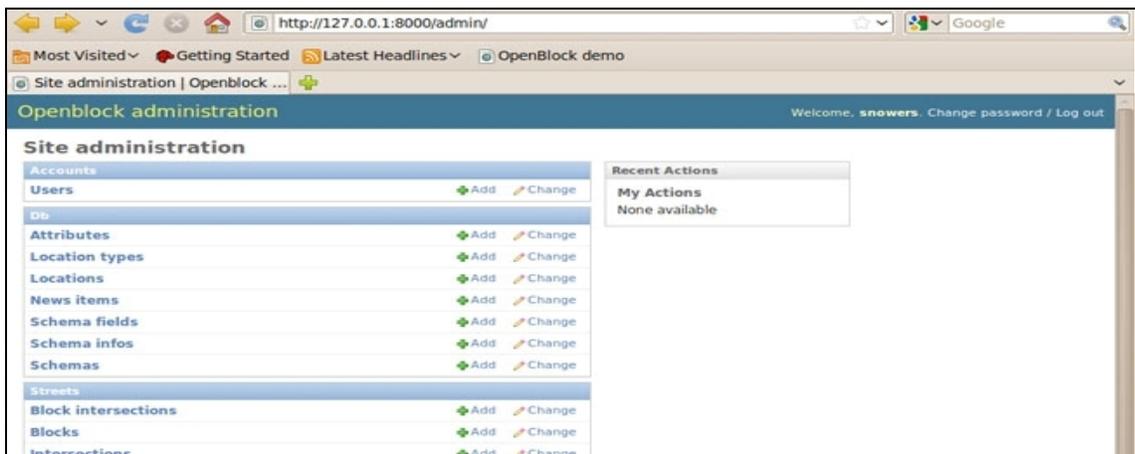


Figure 4: OpenBlock administration page

There are several areas we need to work on to make our Anchorage OpenBlock implementation better and

smoother, including further exploration of OpenBlock's administration utility, and improved user

map search and government data search. We have also started to examine potential uses and opportunities for our block-based open news site once it is up and running. The most important and crucial fact we need to face in the second phase of this project is the sustainability of its application. How to make the site easily accessible and make its content and functionality attractive and essential to the daily life of the general public is our long-term goal. An important aspect of this phase will be to find a suitable business model and determine how to attract the critical mass of regular users needed to sustain this type of local news project.

We recognize that it is not easy to build a new, successful business, even when the idea is good and the market is receptive. Although Lisa Phillips pointed out in her new demographic report on the e-marketer site - The Digital News Audience: 24/7 Participation, "The news may be unpredictable, but its appeal to consumers is not: 92% of consumers use multiple platforms to get news on a typical day, according to the Pew Research Center. These highly engaged consumers see news as a form of social currency and even as a civic obligation. Their need to keep up with current events has them searching out digital news throughout the day, checking in with more than one outlet on more than one device [29]. David Weir also mentioned in his article on bnet.com that the big win for a hyper-local service, theoretically, is local advertising. The media industry is still trying to find a successful online business model. In fact, while we are working on this project The New York Times launched a paid online model. It is too early to judge the success of the NYT experiment but we do not think it is necessary to partner with a large metro daily to create a sustainable local application.

The Anchorage OpenBlock project followed EveryBlock's footsteps and rationale to pitch itself as a "geographic filter" that provides a "news feed" for users' neighborhoods. The planned service provides public records from government databases, news articles from multiple sources, plus blog entries, Flickr photos, Craigslist entries, and user reviews of local businesses via Yelp [30]. Based on our preliminary study and survey, it is our belief that the potential business model we are planning to follow will be leaning toward the recommendation of David Weir described in his article [31] that "the potential business model may need to go with the alternative newsweeklies in each community and subdivision, which have a more granular business model, including stronger relationships with local advertisers, neighborhood by neighborhood, that more closely parallels the hyper-local approach of the original EveryBlock". As mentioned earlier in the introduction,

this will be included in the future study and work of our Anchorage OpenBlock project.

## REFERENCES

- [1] Nicholas Diakopoulos, "Innovation and Computing in Journalism", Journalism and Media Study Center, The University of Hong Kong, <http://jmsc.hku.hk/2011/03/nick-diakopoulos-innovation-and-computing-injournalism/>, March 2011
- [2] Computational Journalism from wikipedia, [http://en.wikipedia.org/wiki/Computational\\_journalism](http://en.wikipedia.org/wiki/Computational_journalism)
- [3] Computer + Journalism site at <http://www.computation-and-journalism.com/main/>
- [4] Anchorage Daily News, "ADN to cut staff by 35; McClatchy drops 1,400," June 16, 2008, <http://www.adn.com/2008/06/16/438029/and-tocut-staff-by-35-mcclatchy.html>, Dec 2010.
- [5] The Associated Press, "10 PERCENT: Daily News will lose 25, half through attrition," September 16, 2008, <http://graphicdesignr.net/papercuts/2008/09/anchorage-daily-news-13/>, Feb 2010.
- [6] Anchorage Daily News, "ADN announces staff, pay cuts", <http://www.adn.com/2009/03/19/728972/adn-announces-staff-pay-cuts.html>, Dec 2010
- [7] EveryBlock.com, "About", <http://www.everyblock.com/about/>, Feb 2010
- [8] The New York Times, "Broad Unemployment Across the U.S.," July 15, 2009, <http://www.nytimes.com/interactive/2009/07/15/business/economy/20090715-leonhardt-graphic.html>, Feb 2010
- [9] The New York Times, "Clean Water Act Violations: The Enforcement Record," September 13, 2009, <http://www.nytimes.com/interactive/2009/09/13/us/0913-water.html>, Feb 2010
- [10] Mindy McAdams, "Flash and data, made for each other", Teaching Online Journalism, September 13, 2009, Feb 2010
- [11] Five Thirty Eight, "Politics Done Right," <http://www.fivethirtyeight.com/>, Feb 2010
- [12] Everyblock, "A News Feed For Your Block," <http://www.everyblock.com/>, Feb 2010
- [13] Knight News Challenge, "Everyblock," <http://www.newschallenge.org/winner/2007/everyblok>, Feb 2010
- [14] Blog.EveryBlock.com, "MSNBC.com acquires EveryBlock", <http://blog.everyblock.com/2009/aug/17/acquisition/>, Feb 2010
- [15] EveryBlock.com, "About", <http://www.everyblock.com/about/>, Feb 2010
- [16] Knight Foundation OpenBlock press release, [http://www.knightfoundation.org/news/press\\_room/knight\\_press\\_releases/detail.dot?id=364383](http://www.knightfoundation.org/news/press_room/knight_press_releases/detail.dot?id=364383), Dec 2010

Stephen Nowers, Yi-Miin Yen, Yoshito Kanamori

- [17] GoogleGroups, “ebcode”, <http://groups.google.com/group/ebcode?lnk=iggc>, Dec 2010.
- [18] EveryBlock.com, “FAQ”, <http://www.everyblock.com/about/faq/#technology>
- [19] EveryBlock.com, “FAQ”, <http://docs.djangoproject.com/en/dev/faq/general/#is-django-a-content-management-system-cms>.
- [20] “GeoDjango Dependencies with Apt-get on Ubuntu”, <http://code.djangoproject.com/wiki/GeoDjangoUbuntuInstall>, Dec 2010
- [21] dkukral, “everyblock”, <https://github.com/dkukral/everyblock>, Dec 2010
- [22] OpenBlockproject, <http://developer.openblockproject.org/wiki>, Dec 2010
- [23] U.S.CensusBureau, [http://www2.census.gov/geo/tiger/TIGER2009/02\\_ALASKA/](http://www2.census.gov/geo/tiger/TIGER2009/02_ALASKA/)), Dec 2010
- [24] Anchorage Daily News lifestyles section RSS feed, <http://www.adn.com/life/index.xml>, Dec 2010
- [25] SeeClickFix RSS feed for Anchorage, <http://www.seeclickfix.com>, Dec 2010
- [26] ebcode, “Source code of EveryBlock.com”, <http://code.google.com/p/ebcode/>, Feb 2010
- [27] “Knight Foundation Expands Neighborhood News Project”, Knight Foundation Press release, [http://www.knightfoundation.org/news/press\\_room/knight\\_press\\_releases/detail.dot?id=364383](http://www.knightfoundation.org/news/press_room/knight_press_releases/detail.dot?id=364383), Dec 2010
- [28] “Quickstart: Install and Set Up Demo Site”, OpenBlock Project <http://openblockproject.org/docs/>, Dec 2010
- [29] LisaPhillips, “The Digital News Audience: 24/7 Participation” <http://www.emarketer.com/blog/index.php/marketing-opportunities-on-digital-news-platforms/>, Dec 2010
- [30] David Weir, “EveryBlock: A New Model for LocalNews?” <http://www.bnet.com/blog/media/everyblock-a-new-model-for-local-news/259>, Aug 2008
- [31] David Weir, “EveryBlock Opens its Code, Seeks aBusinessModel” <http://www.bnet.com/blog/media/everyblock-opens-its-code-seeks-a-business-model/3013>, July 2009