Overview

- **OpenBlock: data centric**
  - Centralization of events
  - Adding unified attributes to data
    - Search by location

- **ScraperWiki: computation centric**
  - Automatic
    - Data analysis
    - Data collection
    - Visualization
    - etc.
A hyper-local news & data platform

- Aggregate all local news onto local maps
- From many sources
  - crime reports, 311 service requests, building permits

Main purposes

- To run a local news website for a community, or
- To create news widgets and feeds for integration with an existing website
  - Quickly geotag articles
  - search by location
Funded by Knight Foundation
  ◦ EveryBlock
  ◦ OpenPlans

Started
  ◦ In June 2009.
  ◦ Created by Adrian Holovaty
Open Block: types of information

- Events
- Police Department reports
- Open311 Service Requests
- Photos from Flickr
  - Meetups
- Local News
  - Restaurant Inspections
Open Block Demo

- **Demo [02127]**
  - News: Hacking
  - Local news
    - Histogram
Weak and strong points

- **Strong**
  - Central news starter
  - Crowed sourced
    - Cheap + high coverage

- **Weak**
  - No videos, audios
    - Video text tags, audio text tags: useful for search
  - No way for professional evaluation
ScraperWiki: Motivations

- A lot of useful data on the Internet
  - Government spending, missing pets ...

- Getting it not easy
  - Different formats: Table, spreadsheet, pdf
  - Different places: scattered over thousand places

- “screen scrapers” used to get @ data
  - Written by programmers
  - But they break, get thrown away, lost ...
An **online** tool
- to make data collection simpler
- collaborative
  - Anyone can write a screen scraper using the online editor.
- Funded by Knight Foundation

**Reusable**
- Like programming libraries
  - The code and data are shared with the world
- Other programmers can contribute to and improve the code.

**Develops a rich toolbox**
- Open source
- Created by thousands of people
Instrument OpenBlock with tools to improve data quality

- At least one user class for journalists
- Classify & rate events and data
- Credible, misunderstanding, etc.

Contribute to ScraperWiki:

- Analysis
- Visualization
- Screen scrapers
  - For websites that don't yet have...