The Sensor Newsroom – A Workshop For Teaching Journalistic Sensing

Fergus Pitt
Senior Fellow, Tow Center For Digital Journalism
Columbia University, New York
ferguspitt@columbia.edu

ABSTRACT
For four weeks from late May 2014, The Tow Center for Digital Journalism at Columbia University ran an intensive workshop to investigate methods of teaching sensor-based journalism. This paper describes the course and its goals, the underpinning theories, details the course design, discusses the most interesting observations and indicates areas for improvement.

KEYWORDS
Sensors, Journalism, Pedagogy

1 - INTRODUCTION

1.1 - Background

The Tow Center for Digital Journalism includes sensors in its areas of focus. That decision is based on three sets of observations. The field of journalism has, over the last ten years, increased its use of computable data as a reporting source, as the raw content for analysis, and built interfaces allowing audiences to explore data[i]. Journalists with the skills to incorporate data into their process are now found in newsrooms. This can be understood as the demand side for sensor data in journalism.

Simultaneously, sensors have become ubiquitous in the developed world. This is the supply side for sensor data in journalism. They are embedded in public and private infrastructure. Sensors are found in many of the finished electronic devices marketed for consumer and industrial users. Electronic component manufacturers, distributors and communities of knowledge form an ecosystem that enables hobbyist prototyping and custom electronic manufacturing as well as niche product development and production.

Given that sensors are a way of observing the world, and reporting observations of the world is an integral element of journalism, it follows that sensors can be journalistic tools. Indeed, a number of journalism projects and ongoing news practices already use sensors, and data originally produced by sensors. Weather reports are a common example of how sensor data are used by news companies. However, sensor-based reporting on single topics include *The Sun Sentinel*'s Pulitzer Prize winning series about illegally speeding off-duty police in Florida[ii], and *USA Today*'s investigations into lead contamination of soil[iii]. WNYC & RadioLab ran a radio and online campaign emphasizing public engagement and education that encouraged their community to build DIY temperature monitors based on the Arduino electronics prototyping platform[iv].

1.2 Knowledge Sources

Although this is the first time a course has run specifically focused on journalistic sensing, it draws from a number of resources.

The broadest context is discussed in the Tow Center report *'Post Industrial Journalism'*; In Section 3: The Ecosystem, the authors note the value of data, the multiplying sources of data, the contested access to data, and the rise of data journalists in newsrooms[v].

The sensing theory and framework draws on concepts from *The Sensor Technology Handbook*[vi], edited by Jon Wilson, the EPA's *Requirements for Quality Assurance Project Plans*[vii].

The journalistic sensing landscape and process was distilled from the following investigative journalism and citizen science projects: The Houston Chronicle's *In Harm's Way*[viii], USA Today's *Ghost Factories*[ix], WNYC's *Cicada Tracker Project*, Public Laboratory's citizen sensing activity. These projects were also documented in the Tow Center Report *Sensors and Journalism*[x].

That report also included essays that formed the basis of the legal and ethical components. In it, Katy Culver noted that a proposed bill of rights for the Internet of Things advocates that the people whose data is collected have ownership of it[xi]. Josh Stearns emphasized collaborative relationships between journalists and their communities[xii]. Lela Prashad noted that physical sensing should be complemented with other investigative methods[xiii].

1.3 Definition of sensors in journalism.

In the Tow Center's definition of the field of sensors and journalism, we included processes where observations of the physical world were mechanically being turned into computable data. So, that includes journalistic processes where journalists are building sensor tools themselves, where journalists are using commercially available sensor tools and where journalists are accessing data derived from sensors or sensor systems.

The author resists using the label 'sensor journalism', with its implication of a separate field of practice. This workshop emphasized the integration of sensor practices with other traditional and modern journalism practices: including interviews and observation.

We excluded processes where humans produce data directly, whether entering data as part of administrative processes or making observations of the world and recording the information in computable form.

We acknowledge there is a valid argument for including in our scope 'software sensing' - monitoring digital activity (especially seeing that when one drills deep enough one reaches physical phenomena: electrons moving). Definitions of 'Robot Journalism' including those by Nick Diakopolous[xiv] and Arjen van Dalen[xv] do not prescribe the source of the data. However, the decision to exclude this type of observation was partially made in reference to our interest in the effects and implications of sensors' physicality. None-the-less, many of the intellectual tools the course covered may apply to 'software sensing' and 'robot journalism'.

2 – THE COURSE

2.1 - Goals

This project had two goals. The first was to develop the Tow Center's expertise in how to teach the use of sensors in journalism. The second goal was to deliver workshop

participants skills and understanding to use sensors as reporting tools.
The activity in the course, including the reporting, analysis and presentation, was in the service of those two goals, as opposed to being goals themselves. This had effects, which are covered below.

2.2 - Workshop Participants
The workshop participants were recruited from pool of applicants. It was open to people who had graduated from a Columbia University graduate program in 2012, 2013 or 2014, or were currently enrolled in a graduate program, including the dual degree program offered jointly by the university's Graduate School of Journalism and the Computer Science School.
The Tow Center publicized the course through its mailing list, twitter, through flyers posted around the Columbia Journalism School, on the Tow Center website and through the School's career services communication channels.
19 people applied, all of who were graduates or current enrollees of the journalism school. Only two had graduated before 2014. Eight were female, 11 were male.
Positions on the course were offered on merit. Applicants submitted their resumes and a cover letter explaining their interest. Shortlisted applicants attended an interview with a single person (this author). Successful applicants were chosen for their demonstrable skills and knowledge in at least one of the following domains: environmental reporting, city reporting, investigative reporting, data skills, and/or teamwork. Five of the students had learnt basic R or python programming.
The course was not for credit towards a qualification of any kind, although the successful participants received a stipend of $750/week paid on completion of the workshop.
Of the eight successful workshop participants, three were female and five were male. One had graduated in 2012, the rest had graduated in 2014. All were drawn from Columbia Journalism School's MS program (two dual-degree students were shortlisted, but both dropped out before the course started). Three were non-US citizens (one Englander, an Egyptian and a French-German).

2.3 - Workshop Format
The course ran full-time for 19 business days, from the 27th of May to 20th of June, 2014. The participants were scheduled to be working for approximately eight hours/day either in the workshop, researching, reporting or producing. The majority of the time was taken up by contact hours.

2.4 - Staffing & Resources
The Tow Center recruited specialists to lead various components of the course. The Tow Center commissioned 10 customized sensor units, based on Arduino components. The sensor cost, including labor, was $310 per unit. The sensors detected PM10 air particles and noise level, ran on battery power and recorded their data to an SD card. They were slightly larger than a shoebox: 34x18x22cm. The bill of parts and control code is available from github[xvi].

2.5 - Workshop Content
The workshop covered an underlying theory and framework of journalistic sensing, stepped through a reporting process, included some practical skills, and examinations of some legal and ethical issues.

2.5.1 - Theory and Framework Sessions
These were the first sessions of the workshop, covering twelve contact hours over one and a half days. There were four modules within this.
The first module covered the same observations detailed in the background section of this paper, but in more depth: That journalism uses data, that sensors are ubiquitous and becoming easier to work with, and that this produces opportunities for journalistic sensing.
For the second, small, module the participants spent an hour reporting tiny stories with the sensors that were immediately available to them: including the accelerometers, colorimeters and cameras on their smartphones, pieces of paper that absorbed moisture. This was an exercise in making recordings of the world to produce computable data. It gave participants hints about error and subjectivity, and pushed them to think creatively about sampling.
The third module was a landscape review. Participants read case studies of journalistic sensing projects, and sensing projects from adjacent fields including citizen science and the arts. These were then presented and discussed in the workshop, with particular reference to the studied projects' goals, outcomes, tools and processes. The case studies were drawn from the Tow Center report *Sensors and Journalism*[xvii].
The fourth module covered a language to describe the most relevant characteristics of sensing for journalism. It introduced the terms (sensitivity to target, insensitivity to interference, spatial and temporal resolution, accuracy, precision, range, linearity, maturity, ownership, autonomy, directness and operating distance) and asked the students to place examples from the previous module on continuum for each characteristic.
These sessions were led by this author with input from Professor Mark Hansen.

2.5.2 - A Sensor-Based Reporting Process
The sessions in this grouping were held throughout the 19-day workshop. The process was somewhat contrived; we started with the need to teach sensing, as opposed to starting with a story that deserved coverage, then selecting sensing as a useful reporting technique (the process that happens in professional newsrooms).

2.5.2.1 - Select the topic and the news context
We chose the topic of construction sites in New York City and their impact on the local environment. This step was made during the course design, around six months before the students were recruited. We hypothesized that we would be able to collect sensor data on dust and noise levels - two outcomes of building sites that may impact neighbors. In the first 'topic' session, Ben Lesser, an experienced CAR reporter who co-taught this part of the course, articulated fundamentals of an investigative journalism project. That was to formulate a story thesis: Defining who the victims could be, what the problem is, who has oversight or regulatory responsibility.

2.5.2.2 - Researching the topic
The students' first step in the reporting process was to research construction in New York, and the potential impacts and regulation of dust and noise. They found out about the mandate of city government departments (NYC Department of Buildings and NYC Department of Environmental Protection) and the mandate of the EPA. They familiarized themselves with New York's "311" municipal complaints process and the construction permitting process. From NYC's open data portal, they downloaded recent complaint and permitting data to get an overview of the datasets' content. They researched the

relevant laws and interviewed experts in the relevant legal domains.
The laws governing noise say construction sound must be no more than 10db above ambient levels, nor above 85db, at 15 feet or more from the source[xviii].
The laws regarding dust are less definitive. They say there must be mitigation methods that prevents 'excessive' dust but does not define a permitted amount of dust release[xix].

2.5.2.3 - Selecting a subject

Once the students were familiar with the domain, under the primary direction of Ben Lesser and Fergus Pitt, they started the process of selecting a particular building site assessing candidates against a number of criteria. This site would be the 'subject' of their investigation. The criteria were:
a) Likelihood of there being a problem. This was ascertained by finding building sites with the highest number of '311' noise and dust complaints and doing news archive searches on the resulting candidates.
b) Whether it would be possible and practical to report on the site (with particular reference to using sensors). This was ascertained by travel time, what stage of construction the project was at, whether the surrounding environment had high levels of 'background' noise and dust, whether there were places the students would likely be able to place the sensors. This was mostly researched through students' site visits.
c) Importance of the potential subject (the number of people affected and/or the size of the development in physical and monetary terms). This was researched through site visits and inspection of publicly available municipal documents.
The students also brought their subjective 'news values' to the assessment. These tended to encompass the elements listed above, but also considered whether the candidates related to other existing economic and social news narratives, such as demographic changes in the local area and the city.

The group started the process with three potential candidates, drawn from the instructors' previous knowledge of the city and their previous examination of '311' data. However, the "successful" candidate building site was found during the students' analysis of 311 data. It was site in Flushing, Queens with a troubled history: It had attracted large numbers of noise complaints throughout its demolition and early construction stages. The building project was running significantly over time and budget, ownership had changed hands, the proposed building use had changed in an attempt to increase the revenue and it was the subject of debate in the local council and local media.
Preliminary interviews with the neighboring residents revealed discontent, some willingness to host sensors, and the students personal observations included very loud construction noise and lots of dust.
This potential newsworthiness outweighed some practical problems: The site was an hour's train travel from the University. Many of the local residents did not communicate clearly in English. Overhead planes flying to and from LaGuardia Airport passed overhead periodically, producing noise at comparably high levels to the construction.

2.5.2.4 - Planning the reporting

During this step, Ben Lesser and Fergus Pitt led the students through the process of deciding where and how to place the sensors, what other information and content the students would need to collect to contextualize and complement the sensor-derived data.
Professor Mark Hansen contributed to the discussions, with a particular emphasis on the issues of consent from the residents.

The building site in Flushing was neighbored by a four-story residential apartment block on one side, a church-owned three story residential building on the other side, a supermarket at the back, and a minor collector road at the front. On the other side of the road were more apartment blocks of 4-9 stories. The closest wall and windows adjacent apartment block were well within 15 feet of the subject building (i.e., they were too close to be protected by the noise legislation as worded).

The sensor placement considered the following practical and ethical factors: To where the students had legal and physical access, the distances from the site used by the legislation governing noise pollution, whether the sensors could collect information harmful to the residents, where the sensors would collect data that could produce meaningful information.

2.5.2.5 - Doing the reporting

Starting between the seventh and ninth days of the course, the students placed two sensors in an apartment block diagonally across the road (day 11), three in apartments directly adjacent to the building site (day 11), and one more "control" sensor in the grounds of a synagogue 250 meters away (day 11), placed one sensor in the church grounds (day 13). The sensors recorded a data point each second for the voltage coming from the microphone (which varied in proportion to noise volume), and a count of PM10 dust particles.

The students verbally briefed the sensor-hosting residents on the workshop's aims and the sensors' operations. The residents also took one-page consent form that also had a textual explanation of the project (written only in English). The residents who hosted the sensors were also asked to fill out diaries documenting their observations of the noise and any dust with precise times. One resident did so.

The first batch of data, from the first six sensors, was collected on day 13.
The second batch of data was collected on day 15 from six sensors, and the final sensor and data was collected on day 16 (which also included weekend days' data).

When the data was collected, the students did a quick time-series visualizations of each unit's recorded data and explained what that meant to the host residents, before again asking their permission to take the data.

The data and, where available, the resident's observation diary were also used as the basis for interviews with the residents where the students worked to elicit the residents' impressions of the noise and dust and other aspects of living near the building site.

The students also performed more 'traditional' reporting to gather content that would be needed in a publishable news piece: They gathered more information about the building site's history, photographs of the site and the residents, audio recordings of the site noise, and interviews with experts in health and construction.

2.5.2.6 Data Interpretation

The students learned some principles of data analysis, using iPython notebook[xx] running on a large Amazon EC2

instance. These sessions were led by Mike Dewar, a data scientist who works at the NYT R&D lab.
The sessions covered a) the concepts and production of time series to show and compare change over time, and a histogram to show and compare distributions, b) two approaches to interpreting noisy raw datasets; a moving average and finding 'unnaturally shaped' spikes, and c) some basic principles of sampling design. They used the SciPy collection of Python packages[xxi]. The sessions took one and a half days. The first, full, day was delivered before any field data had been collected (it worked with data from tests of the sensor units), the second half day used the first batch of data from the field.

2.5.2.7 Presentation

The final part of the course's journalistic sensing process concerned presentation and user experience. Kio Stark, the senior producer of WNYC's data journalism team led a workshop exploring alternative ways of presenting and experiencing digital content. Ben Lesser returned to lead the writing of a more traditional written piece.

2.5.3 - Technical Skills

Throughout the course the students were introduced to some specific technical skills for journalistic sensing.

2.5.3.1 - Electronics

The students spent one and a half days learning the fundamentals of electronics and prototyping. These sessions were led by Kipp Bradford and Julie Steele, who build custom electronic sensing installations for conferences and other businesses.
Through the process of building an Arduino-based local wireless temperature and humidity sensing network, the sessions covered the concepts of voltage, current, resistance, conductors and insulators, and gave the students experience breadboarding and soldering.

2.5.3.2 - Remote Sensing

The students did a one-day session introducing remote sensing. Lela Prashad, a data science consultant who also works with NASA, led the sessions.
Prashad covered the underlying physics of remote sensing, the data/image formats, briefed the students on the different government and private satellites, showed them how to find and download data and do basic moisture-level interpretations.

2.5.3.3 - Data Interpretation

See 2.5.2.6

2.5.3.4 - Drone Piloting

The students did a short session learning to pilot a consumer-grade UAV; the DJI Phantom2.

2.5.4 - Legal and Ethical Considerations of Sensors and Journalists

These topics were covered in the specific sessions below, and also incorporated into the reporting planning, reporting, data interpretation and presentation processes. The content was grounded in the essays contained in the laws and ethics section of the Tow Center report *Sensors and Journalism*[xxii].

2.5.4.1 - Laws

This half-day session was framed as 'How a newsroom lawyer would negotiate this workshop's sensing process'. It was led by Deirdre Sullivan, a senior counsel for the New York Times Company, who was at pains to point out she was not expressing company policy.
Sullivan started by noting that legal considerations are very specific to the facts of the individual case, and did not try to cover all the potentially relevant law for sensors and journalism. Considering the workshop's project, she identified potential main issues of personal harm and negligence, the contractual relationship between journalists and equipment suppliers, trespass, invasion of privacy and defamation. She talked about how she would try to negotiate and mitigate each of these issues.

2.5.4.2 - Ethics

This half-day session focused mainly on the relationships between journalists and the people they report on. Josh Stearns, Director, Journalism & Sustainability for the Geraldine R. Dodge Foundation led the session.
The key concepts discussed were privacy, transparency, trust, objectivity, data ownership and data access. The most heavily examined ethical issue, in this session and during the reporting planning process, was that of the interaction between journalists and the communities where the sensors were placed.

3 – OBSERVATIONS & ASSESSMENTS

The students filled out course evaluation forms, rating each topic out of five and making free-text comments. They also participated in a group debrief discussion on the last day of the workshop. This section is drawn from those resources, from observations during the workshop and from discussions with the session instructors after the workshop.

The students' course evaluations were complementary (although we have no baseline with which to compare). Within those positive ratings, the students rated highest the data analysis (2.5.2.6), the electronics (2.5.3.1), the legal and ethical (2.5.4), and the sensing context sessions (2.5.1). The reporting planning sessions (2.5.2.4), the collection and interpretation and the writing sessions were rated lowest (part of 2.5.2.7). No student rated anything below a 3/5.

The theory and framework were particularly successful, from this author's observations. Throughout the rest of the course the students kept using the language these sessions introduced and compared their work to the work in the case studies.

When researching the reporting topic; construction in NYC, the students struggled to record the regulations with sufficient detail and precision to be compared to sensor-derived data. Initially they could not answer questions about noise units or the distances where observations would need to be recorded. This may reveal an important distinction between reporting skills and processes that focus on qualitative experiences of interviewees and quantitative data collection by sensing equipment. Sensing for investigative journalism requires high precision, both in the data collected and in journalists' understanding of the laws or health effects to which the data is being compared.

The course design favored exploration and participation over the reliability of the produced data: The course worked with extremely immature sensing equipment. The data collection process was planned very quickly, and had to deal with the complexity of an urban neighborhood. The students were learning wholly new skills and working on unfamiliar subject matters. An alternative would have been to buy mature sensing equipment, record data in much more controlled circumstances. The cost was not producing conclusive data, but the benefit was having the students gain familiarity with more parts of the data-production process. It is not clear whether that design decision is correct in all circumstances.

The custom-built sensors suffered a number of faults over the period. Out of ten sensors delivered, the time-stamping function failed on three of the sensors. One sensor would not accept re-programming, and power failures rendered two inoperable.
Although faults are common in immature equipment[xxiii], the author's procurement process left room for improvement.

The timing of the course introduced challenges. It was run over 19 consecutive business days. This limited the time the sensors could be in the field collecting data, and allowed very little time to make changes to the sensing hardware or respond to the abilities of the students or new information about the subject of the story.

The 'data interview' noted in 2.5.2.5 may be an interesting new reporting form. The practice has two benefits: Residents are informed and have more control over the data produced by the provided sensors, and the data becomes a prompt for the interviewee's impressions and interpretation.
This form has echoes of other collaborative or 'UGC' journalistic processes: a number of news and information organizations provide 'citizens' tools to produce their own content; from cameras to writing platforms[xxiv]. News organizations also coach 'citizens' in various journalistic and media production skills.
However, the practice faces headwinds. In our experience some of the students reported that the residents were not particularly interested in discussing the data in detail. Commentators on participatory journalism have observed that participants' motivation to engage relies on them receiving clear benefit that obviously outweighs the cost and beats whatever else in their lives competes for their time, effort and attention[xxv].

4- CONCLUSION

This workshop provided value in two areas. As indicated by the observations in section 3, many potential adjustments exist. However, the student/participants evaluated the course highly and appeared to gain new skills and knowledge. The course also provides a 'first draft' of how journalistic sensing might be taught. Future iterations of the syllabus can incorporate lessons from the observations and be adapted to new settings.

---

[i] Cohen, Sarah, et al. "Computational Journalism: A Call to Arms to Database Researchers." *CIDR*. Vol. 2011. 2011.
[ii] Kestin, Sally, John Maines, and Sun Sentinel. "Cops among Florida's Worst Speeders, Sun Sentinel Investigation Finds." *Sun Sentinel*. Accessed August 31, 2014. http://www.sun-sentinel.com/news/local/speeding-cops/fl-speeding-cops-20120211,0,3706919.story.
[iii] Young, A. USA TODAY. "Ghost Factories: Smelting and Lead Contamination." *USATODAY.COM*. Accessed August 29, 2014. http://www.usatoday.com/news/nation/smelting-lead-contamination.
[iv] "The Cicadas Are Coming: WNYC's Tracker Is the Latest Sign of the Rise of Sensor News Networks." *Nieman Lab*. Accessed August 31, 2014. http://www.niemanlab.org/2013/03/the-cicadas-are-coming-wnycs-tracker-is-the-latest-sign-of-the-rise-of-sensor-news-networks/.
[v] Anderson, Christopher William, Emily Bell, and Clay Shirky. *Post-industrial Journalism: Adapting to the Present: a Report*. Columbia Journalism School, 2012.
[vi] Wilson, Jon S. *Sensor technology handbook*. Elsevier, 2004.
[vii] "EPA Requirements for Quality Assurance Project Plans." Environmental Protection Agency, March 2001. http://www.epa.gov/QUALITY/qs-docs/r5-final.pdf.
[viii] "Chronicle Cross-County Study Reveals Risky Load of 'Air Toxics.'" *Houston Chronicle*. Accessed August 29, 2014. http://www.chron.com/news/article/Chronicle-cross-county-study-reveals-risky-load-1643020.php.
[ix] Young, A (ibid)
[x] Pitt, Fergus. *Sensors and Journalism*. Tow Center For Digital Journalism, Columbia University, May 30, 2014. http://towcenter.org/sensors-and-journalism/.
[xi] Culver, Kathleen. "Rights, Ownership, Ethics and Sensors in Journalism." In *Sensors and Journalism*, edited by Fergus Pitt, n.d. http://towcenter.org/sensors-and-journalism-rights-ownership-ethics-and-sensors-in-journalism/.
[xii] Stearns, Josh. "The Ethics of Sensor Journalism: Community, Privacy, and Control." edited by Fergus Pitt, n.d. http://towcenter.org/sensors-and-journalism-the-ethics-of-sensor-journalism-community-privacy-and-control/.
[xiii] Prashad, Lela. "The People Within The Pixels." In *Sensors and Journalism*, n.d. http://towcenter.org/sensors-and-journalism-the-people-within-the-pixels/.
[xiv] Diakopoulos, Nick. "The Anatomy of a Robot Journalist." Accessed September 1, 2014. http://www.nickdiakopoulos.com/2014/06/12/the-anatomy-of-a-robot-journalist/.
[xv] Van Dalen, Arjen. "The Algorithms Behind the Headlines." *Journalism Practice* 6, no. 5–6 (March 30, 2012): 648–58. doi:10.1080/17512786.2012.667268.
[xvi] "SensorNewsRoom - Code and Bill of Parts Repository." Accessed September 1, 2014. https://github.com/ferguspitt/SensorNewsRoom.
[xvii] Pitt, F, (2014) ibid
[xviii] Local Laws of The City Of New York, 2005, New York City Administrative Code §24-202 http://www.nyc.gov/html/dep/pdf/law05113.pdf
[xix] Local Laws of The City Of New York, 2005, New York City Administrative Code – Code §24-146(c) http://www.nyc.gov/html/dep/pdf/dust_rules.pdf
[xx] Fernando Pérez, Brian E. Granger, *IPython: A System for Interactive Scientific Computing*, Computing in Science and Engineering, vol. 9, no. 3, pp. 21-29, May/June 2007, doi:10.1109/MCSE.2007.53. URL: http://ipython.org
[xxi] *SciPy - Scientific Computing Tools for Python*, n.d. http://www.scipy.org/about.html.
[xxii] Pitt, F, (2014) ibid
[xxiii] Sharma, A., Golubchik, L., Govindan, R. 2007. 'On the prevalence of sensor faults in real-world deployments'. In *Proceedings of the IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*.
[xxiv] Hermida, Alfred, and Neil Thurman. "A Clash of Cultures." *Journalism Practice* 2, no. 3 (September 2, 2008): 343–56. doi:10.1080/17512780802054538.
[xxv] Daugherty, Terry, Matthew S. Eastin, and Laura Bright. "Exploring Consumer Motivations for Creating User-Generated Content." *Journal of Interactive Advertising* 8, no. 2 (March 1, 2008): 16–25. doi:10.1080/15252019.2008.10722139.