Array of Things: A Scientific Research Instrument in the Public Way

Charles E. Catlett
University of Chicago & Argonne National Laboratory
Chicago, Illinois
catlett@anl.gov

Peter H. Beckman
Northwestern University & Argonne National Laboratory
Chicago, Illinois
beckman@anl.gov

Rajesh Sankaran
Argonne National Laboratory
Chicago, Illinois
rajesh@anl.gov

Kate Kusiak Galvin
University of Chicago
Chicago, Illinois
katekusiak@uchicago.edu

ABSTRACT

The "Array of Things" (AoT) project aims to create an urban-scale instrument for research and development across many disciplines. The concept is to exploit Internet of Things (IoT) technologies to build an instrument analogous to an array telescope, where many identical detectors spread over an area work as a unit. AoT, then, is an IoT-enabled "telescope" pointed at the city. With support from the National Science Foundation, the University of Chicago, Argonne National Laboratory, the City of Chicago, and industry the project has adapted an Argonne-developed resilient sensor-hosting platform, Waggle, for urban installations. The project will install 500 units, or "nodes," by late 2018, with installation in phases to allow for technology improvements based on evaluation of early installations as well as to enable one or more insertion points for component upgrades and expansions, such as emerging sensors. This paper describes the initial stages of the project, focusing on lessons learned in areas ranging from resilient technical design to manufacturing to privacy policies and public engagement.

CCS CONCEPTS

• **Hardware → Sensor applications and deployments**; • **Computer systems organization** → Sensors and actuators; Sensor Networks;

KEYWORDS

wireless sensor networks, distributed systems, data dissemination, edge computing, urban sensing

ACM Reference format:
Charles E. Catlett, Peter H. Beckman, Rajesh Sankaran, and Kate Kusiak Galvin. 2017. Array of Things: A Scientific Research Instrument in the Public Way. In *Proceedings of The 2nd Workshop on Science of Smart City Operations and Platforms Engineering, Pittsburgh, PA USA, 21 April 2017 (SCOPE 2017),* 9 pages.
DOI: 10.1145/3063386.3063771

SCOPE 2017, Pittsburgh, PA USA

DOI: 10.1145/3063386.3063771

1 INTRODUCTION

Unprecedented expansion of infrastructure to support urbanization around the globe represents an opportunity to improve human well-being and productivity, reduce human impact on the planet, and increase the resiliency and sustainability of cities. Yet many such opportunities remain unrealized despite ongoing and significant urban challenges ranging from energy to transportation, from water to air pollution, and from education to crime to healthcare. These factors suggest a need to enable interdisciplinary teams to better understand urban infrastructure and systems, urban impact on the natural environment, and the interplay between urban spaces and human behavior. Achieving this goal will require data [27], including instrumenting cities with new types of sensors and measurement strategies.

Concurrently, cities are increasingly characterized by powerful and ubiquitous mobile and embedded technologies, including new capabilities based on sensing, artificial intelligence, and information systems interactions, with promises such as reduced road congestion and net-zero-energy buildings. Research in architectures and technologies necessary to create such embedded technological capabilities and to understand their impact on infrastructure, economic, social, and behavioral dynamics will require urban-scale prototyping and experimentation.

AoT, then, is designed to support three general classes of research: (a) analysis of data, providing urban measurements with greater spatial and temporal resolution; (b) research and development of edge computing capabilities; and (c) research and development of new embedded hardware systems—such as new computing, sensing, or communications devices.

To develop an instrument that would support such a research agenda, the Array of Things (AoT) project began with a series of scientific workshops engaging various science communities, often beginning with questions such as, "for your research, if you could place technology throughout a major city—sensors, cameras, microphones, computation—what would you want to install?" The AoT team's experience placing experimental sensors in Chicago suggested that full implementation of a persistent scientific instrument would require partnership with city government and, specifically, with units that install and manage public infrastructure. Installing persistent infrastructure is also expensive relative to the price of a typical sensor device, so that device would need to deliver value over a long period of time. The instrument design then would need to balance the need for a reliable platform [15, 22] with the need for enabling the research community to obtain useful measurements and to contribute new hardware and software. Argonne's Waggle platform, designed to provide reliable sensing capabilities in remote locations [5], provided the basis for AoT's urban installation. In the following sections of this paper we discuss the resulting platform design process, early manufacturing and installation experience, and next steps.

We begin in Section 2 with the scientific requirements of the instrument, outlining the process used to engage over 50 scientists from dozens of universities, national laboratories, and companies and from as many different disciplines.

Section refmeasurements translates these science requirements into design principles for the platform architecture to support the three use modalities, (a) measurements, (b) edge computing, and (c) new hardware, as well as implementation requirements including (d) scalability, (e) high reliability, and (f) replicability. Section 4 deals with design and packaging to support public installation, considering future upgrades and the need to protect electronics while also exposing sensors to the environment.

In section 5 we provide an overview of the AoT data pipeline necessary to make data and metadata available in various forms to scientists and to the general public. Section 6 outlines the process of engaging the public to develop policies and partnerships guiding the operation of a scientific instrument in the public way. Section 7 reports on early results of manufacturing and installing devices in Chicago, and section 8 outlines next steps and long-term vision.

2 SCIENCE-DRIVEN FUNCTIONALITY

To explore the notion of an urban sensing infrastructure, the AoT team held a series of scientific meetings and workshops beginning with a workshop focused on urban climate, hosted at Argonne in 2013 [14], two general urban sensing meetings with scientists from a number of universities and national laboratories in 2014, and an "instrumented cities" workshop with scientists from the U.S. and U.K. in March 2015 [9]. The team also engaged social, behavioral, and economics research communities through the NSF-supported Urban Sciences Research Coordination Network (US-RCN, [10]), including workshops in 2013 and 2014 .

These interactions revealed the need for three "use modalities" for an urban-scale instrument. The first considers AoT as a traditional scientific instrument producing *data*. For these users, AoT will provide high spatial and temporal resolution sensor data from an active urban environment for analysis and integration into scientific workflows, mobile applications, or other services. The second modality involves using AoT as a secure *edge computing* platform for embedded information services or knowledge gathering using in-situ processing of sensor data to develop new and/or integrated data products, such as the use of machine learning techniques to extract features from imagery or sound. The third modality considers AoT as a platform for *testing new sensors, communication devices, computational devices*, or other hardware embedded in a major urban area.

Sensors in the External Stevenson Shield	
Waggle MetSense:	Acceleration/vibration, barometric pressure, humidity, light, temperature.
Waggle ChemSense:	CO, H_2S, NO_2, O_3, SO_2
Individual Packaged Sensors:	RMS Sound Level, particles, camera
Sensors in the waterproof enclosure	
Waggle LightSense:	Internal temperature, humidity, magnetic field. External light, IR, UV
Individual Packaged Sensors:	Camera (upward facing)

Table 1: Sensors in 2016 AoT Configuration

2.1 Measurements

Traditional air quality and weather monitoring has required carefully calibrated sensor stations that are expensive to establish and operate, resulting in sparse networks across major cities and significant gaps in coverage. Consumer weather stations have filled in some gaps in weather data, but air quality remains expensive to measure with the level of accuracy needed to support scientific analysis. The temporal and spatial dynamics of air quality across any city are significant enough that sparse networks at best provide general information about the urban region. These stations also often show different measurements even when located close together [30].

Scientists investigating air quality, weather, and climate also expressed interest in sky images, and those studying energy demand and energy efficient building design and building controls suggested measuring ambient light, ultraviolet (UV), and infrared (IR), which would be useful in estimating solar load on nearby buildings. These discussions led to the inclusion of an upward-facing camera and light, UV, and IR sensors (LightSense in Table 1).

Creating a platform to provide consistent measurements, from hundreds of devices—with the accuracy and sensitivity required for scientific investigation—presents challenges related to calibration and quality control that are less evident (or perhaps less discussed) in low-cost sensing projects.

2.2 Edge Computing

Support for edge computing—embedding intelligence in the devices for in-situ processing—was initially motivated by the need for image analysis. Community organizations expressed interest in automated flood detection, and transportation researchers needed sequences of images to understand the flow of vehicles and pedestrians through an intersection. Traditionally, this type of image or video analysis requires collecting images or video continuously and analyzing it centrally, in the cloud. However, this approach introduces at least three significant challenges. First, the communications costs would be substantial in any urban area lacking free high-speed networking. Second, collecting and storing such information would introduce privacy concerns given the images would most certainly include faces, license plates, and events such as crimes or traffic accidents, subjecting the project to disclosure requirements and costs. Third, centralized processing also introduces centralized failure points—a resilient urban instrument should not fail if the cloud service is unavailable. Here the concept of edge computing offers a way to support image analysis without incurring the costs of communications or the privacy risks of compiling exhaustive image archives. It also supports resilience in that nodes can detect and notify independently. Finally, the use of edge computing for image and audio processing provides mechanisms for very powerful and demonstrable privacy protections [29]. (We discuss privacy in more detail in § 6.)

Social scientists were also enthusiastic about edge computing. Observational data is central to understanding the use of public spaces [28] and to characterizing urban environments [24], but is expensive to obtain. Consequently, data about the use of public spaces is often anecdotal. Pedestrian and vehicle flow is similarly estimated and/or sampled. For example, traffic flow for the Chicago Loop is estimated based on the GPS-tracked movement of public buses, while vehicle count is based on observational surveys done every 10 years [12]. The inclusion of cameras, a microphone, and edge computing introduces opportunities for social scientists and machine learning experts to collaborate to create new forms of "automated observations" examining street activities such as the average size of groups using a park or the flow of pedestrians through an intersection.

Computer scientists also expressed interest in edge computing as a way to support research into new forms of hyper-local computing [25] or information services, or for investigating new embedded urban systems concepts such as vehicle-to-infrastructure communications and services.

2.3 Hosting Guest Sensors

A final use modality is to test new sensors and devices in an urban area. Sensor technology is rapidly advancing. Microfluidics-based, MEMS, and screen printed electrochemical gas sensors [6] are steadily replacing more traditional sensors, often with significantly lower power and cost. These sensors are can measure airborne particles [21] and a wide range of pollutants. AoT nodes have been designed with an extensible communication architecture to support additional sensors via expansion of the sensor enclosure. Furthermore, computer scientists and engineers proposed experiments including cosmic ray detectors, software defined radios, and experiments in wireless mesh networking.

3 PLATFORM ARCHITECTURE: WAGGLE

Waggle [4], the Argonne-developed resilience sensor platform, comprises remotely installed devices ("nodes") that communicate with a central data and management service ("Beehive"), including a software stack that runs on the nodes and a software stack that defines Beehive, an integrated set of virtual servers (Fig. 1).

The platform supports the use modalities discussed above—(a) measurements, (b) edge computing, and (c) device hosting)—while optimizing for (d) extensibility and ability to scale to thousands of nodes, (e) high reliability and resilience, and (f) replicability. Underlying all of these requirements was the importance of open systems

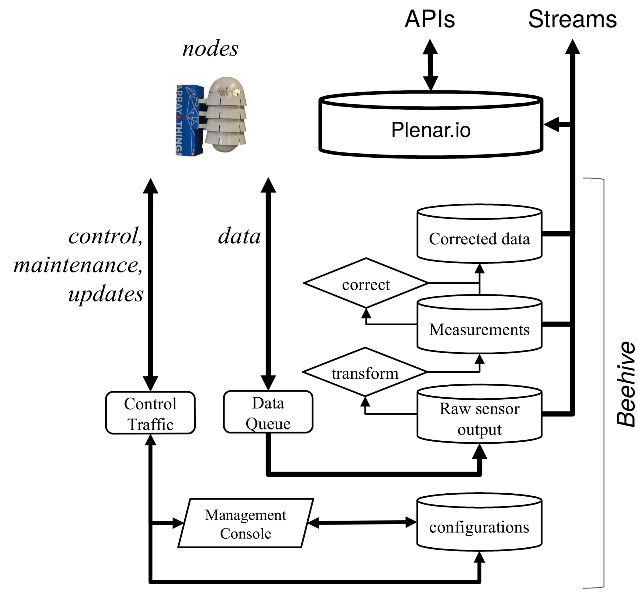

Figure 1: Beehive manages control and data for aggregates of nodes.

to support the software environments used by the research community (Linux) and standard hardware interfaces for supporting guest sensors.

3.1 Measurements

As noted above, scientists requested a wide range of sensors to be included in the AoT instrument. To support their science, however, it was essential for the platform to capture the context in which each measurement is made. Scientists participating in Argonne's urban climate workshop [14] stressed the need for detailed context in order to properly interpret a given measurement. To illustrate, a temperature reading in an area with direct sunlight will vary widely from a temperature reading only a few feet away in the shade.

AoT nodes include a sensor board that hosts common environmental sensors (MetSense in Table 1), a sky-facing sensor board (LightSense in Table 1), and an experimental board with novel printed electrochemical gas sensors (ChemSense in Table 1).

3.2 Edge Computing

Support for standard libraries and packages is essential for the research community to readily adapt their code to an instrument. The edge computing community relies heavily on Linux-based environments, libraries such as OpenCV [20] and OpenCL [23], and frameworks including TensorFlow [1] and Caffe [17]. There are also large developer communities using Android or Java-based platforms, but the machine learning research community relies most heavily on Linux. From a platform standpoint, the decision to begin with a Linux edge computing environment was based on input from the developer community and on available open source software. In the future, as computational hardware to support machine learning begins to become available, alternative edge computing platforms may be introduced into the Waggle platform.

3.3 Hosting Guest Sensors

The initial versions of Waggle comprised several core components, including external communications, security protocols, sensors (Table 1), on-board data caching and node management, data movement libraries, and edge computing libraries. The concept that AoT also needed Waggle to support extensibility—for students and professors to develop and test new sensors and computer vision algorithms—emerged from the workshops and interactions with scientists interested in urban and environmental sensing. While several modular sensor platforms existed, these were limited to modules already conceived of and provided by the vendor. For such platforms, "modular" meant that scientists could field a subset of the available sensors, but the platform was not extensible.

Waggle employs several approaches to hosting sensors. Standalone sensors with sophisticated, embedded, on-board processing, such as hyperspectral imagers, can be integrated via standard network technologies (WiFi, Ethernet). Sensors with simple control interfaces, such as the Alphasense particle sensor, can be connected by USB, I^2C, SPI, or simple serial. Waggle can also support simple analog sensors via the MetSense board, which includes an Arduino-class microcontroller for interrogating sensors.

Each of these hosting strategies involves a different set of design and integration efforts, with different lead-time requirements. New sensor devices that require space in the Stevenson radiation shield are limited to roughly three inch in diameter and one inch in height, and their integration requires modification of the sensor harness and adding layers to the shield to extend the internal cavity. Connecting sensors to the MetSense board requires both wiring to empty circuit board headers and modifications to the firmware. Each of these integration efforts also involves a different set of engineering, cost, and lead-time considerations. For AoT, we expect that such extensions will happen annually. However, the platform is available for third parties to modify as well.

3.4 Scalability and Extensibility

As with the modular node hardware architecture, the data pipeline uses a plug-in software architecture implemented on a Linux single-board computer (SBC). To add a new sensor, a client-side plug-in runs on the node to read sensor output and push it to the node software layers responsible for the data pipeline. These layers cache data locally and then pass the data over a secure encrypted communications link to Beehive. The local data cache is large enough to hold several months' worth of data, and the data pipeline removes data from the cache only after receipt acknowledgement from Beehive. Data from all sensors is transmitted at 30-second intervals.

At Beehive, a software plug-in decodes the sensor data, applies corrections (if any), and converts from native sensor output to units of measure (e.g., degrees Celsius or parts-per-billion). Every sensor is hence implemented with three components—sensor hardware, node-side plug-in and driver, and Beehive-side plug-in.

Beehive also provides node configuration and management functions, acting conceptually as a head node for a parallel computer or cluster. The primary difference is that the Waggle nodes are physically remote and predominantly communicate through the public Internet, thus communications involve much higher latency and lower bandwidth than a network in a rack or data center. In this context, managing different configurations on subsets of nodes—partitions—can be accomplished with well-established techniques, tools, and frameworks. Partitioning will be important for (a) testing software configurations, including virtual containers for edge computing, (b) accommodating different components associated with multiple generations of nodes, and (c) unique configurations, operating parameters, and policies for installations in different cities.

We anticipated (a) and (b) early in the project. The AoT plan is to deploy nodes in multiple phases, allowing upgrades and improvements in between; the 500 Chicago nodes will comprise 2 to 3 generations. These will require modified software stacks (e.g., drivers and plug-ins for new sensors). Further, it will be useful to allow for experimental partitions to support the research community developing node software.

We did not initially anticipate installations in multiple cities, but to date over 75 cities and universities have expressed interest. Supporting (c) became important given that these different deployments may have different hardware and software configurations, communications technologies, and/or operating parameters and policies.

3.5 High Reliability

Waggle's original design contemplated aggregates of nodes installed in remote locations, where physical access is cost- and/or time-prohibitive. Further, in such locations, the installation costs would be high relative to node cost, suggesting that mean-time-to-failure should be measured in years. Simply put, the design goal involved adopting design principles used for satellites or deep space probes [2].

Typical sensor networks are simple—a sensor device, a microprocessor to read the sensor(s), and a modem to transmit data. Mature, reliable, robust technologies exist for these components, assuming that there are reliable sensors for the desired measurements. But the design goal of embedding experimental computation and sensors, including high performance remote programmable edge computing systems often with comparatively brittle software stacks, introduces challenges to reliability.

At the heart of the Waggle node, then is a supervisor subsystem—Waggle Manager (WagMan). WagMan's role is to ensure reliability not only in context of hosting programmable (subject to reliability issues, self-imposed or imposed by programmers) computers, but also in outdoor environments where heat, moisture, dust, and other factors affect reliability. WagMan comprises the following subsystems:

- a control microprocessor,
- real time clock,
- sensors to monitor the Linux SBCs (temperature, current draw, digital heartbeat),
- enclosure internal temperature and humidity sensors,
- electronic switches to change the boot medium selection for the Linux computers, and

- relays and USB interface to support hard and soft reset of the Linux computers and other hosted payloads (such as those shown in Table 1).

WagMan's microprocessor monitors heartbeat signals, current draw, and CPU temperature to track the health of the Linux SBCs. WagMan also monitors temperature and humidity within the node enclosure, and can signal the Linux SBCs to shut down if the environmental parameters are outside of their safe operating ranges. When the environment returns to within the working limits of the devices, WagMan can reboot them. Each node also includes a Linux SBC to support node subsystems, data management, security protocols, etc. A second Linux SBC with GPU cores supports edge computing such as machine learning and image and sound processing.

Selection of the Linux SBCs also considered reliability, and the AoT nodes use Odroid [16] SBCs. A key factor in this decision was the support for two separate storage devices—a micro-SD card, and an embedded Multi-Media Controller (eMMC) card, along with the availability of GPIO ports for implementing heartbeat. WagMan tracks any boot failures on the SBCs, switching first to the recovery boot device in case of multiple successive failures or instability of the primary boot device. The primary boot device can then be accessed, recovered if possible, and remotely repaired from Beehive. In case of irrecoverable hardware damage to the primary boot device, the Linux computer is able to continue to function by reconfiguring the recovery boot device as the primary boot device. Hence, the dual storage device feature both improves node resilience, and also extends its lifetime with acceptable graceful performance degradation. WagMan itself is field-upgradable in that the Linux SBCs can update its firmware. WagMan also has built-in safety features for recovery from faulty flashing process or power failures during flashing.

3.6 Replicability

The majority of scientists involved in planning AoT were interested in using measured data, and this also true of the over 75 requests for information about installing AoT in other cities. Although some groups plan to install a full instance of Beehive and a collection of nodes, the vast majority have interest in siting nodes, but not in configuring, managing, and operating a Beehive. For this reason, pilot projects are being implemented as turnkey systems, with nodes operating as appliances that are centrally managed from Chicago. As with AoT in Chicago, data from these installations will be openly available through Plenario [11], an open source AWS-cloud hosted geospatial data search and exploration platform designed to support open data search and examination developed by UChicago. The appliance model ultimately leverages Waggle's scalability, reducing support costs and eliminating the need for technical staffing at each pilot installation.

4 DESIGN FOR URBAN INSTALLATIONS

There were several important factors related to the physical packaging and design of the AoT nodes. The components, including all of the sensors requested from the science community (Table 1), needed to be packaged in a form that could be securely mounted in the public way, exposed to the elements. To emphasize the concept of AoT as a public utility, the design criteria also included aesthetic considerations—the devices should be brightly colored to be easily seen.

Prior to the science community interactions, partners at the School of the Art Institute of Chicago (SAIC) taught a graduate course oriented around the design of an AoT node, with criteria balancing function and "engaging" form. Students prototyped several minimalist approaches, including simple non-Faraday enclosures with power, Internet, and standard interfaces that any group could use to deploy a device. These approaches may have met the needs of the hosted device use modality discussed earlier, but the science community with interest in measurement data was actually the largest, and these minimalist methods would not address their needs. We also considered the notion of creating a Waggle node design that would be field-upgradable with new hardware components, but determined that the cost of swapping a unit would be lower than the cost of sending a trained individual to open, modify, and test the node. Swapping in a lab-tested unit would also be a more reliable approach.

The first enclosure (Fig. 2-L) was a single housing, with external AC/DC power supply and a vented section for sensors. The single enclosure made it more challenging to protect the electronics from moisture (and insects), and also restricted airflow. Engineering for the addition of two cameras also required a reconsideration of the packaging. The current design (Fig. 2-R) places the sensors in a Stevenson radiation shield, separate from the sealed, waterproof electronics enclosure. Light, UV, and IR sensors are housed at the top of the electronics enclosure, as these do not require external airflow.

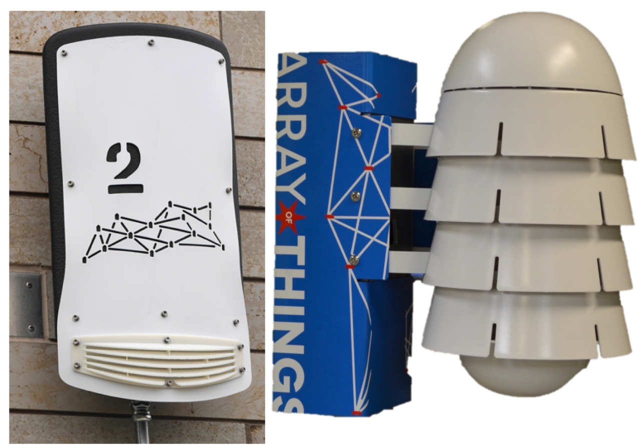

Figure 2: Early design installed at UChicago (L) and current design (R).

In addition to providing installation labor as in-kind support, the City of Chicago is also providing electricity. Nodes are hardwired, drawing under 50 watts. We determined that wired power would enable future upgrades in computing power while also making the installation more streamlined than one with a solar panel. Through collaboration with city electricians, we also adopted standard mounting hardware and designed the units to minimize installation or swap time, reducing cost of installation and of future upgrades. For safety reasons and to simplify design (including a low barrier for adding hardware modules), there are separate waterproof enclosures for AC and DC components.

Early in the project we considered mounting the devices 5-8 feet above grade, but decided to mount at 20-22 feet for two reasons. First was a concern regarding vandalism or theft. Second, a higher installation allows for a larger field of view for the camera, allowing images of the entire intersection.

5 DATA ARCHITECTURE

Environmental scientists noted that sensor response can drift over time, requiring re-calibration. The number of nodes, their remote locations, and physical access constraints make typical re-calibration and maintenance procedures expensive. Therefore in AoT any drift will be addressed through corrections in the measurements. These corrections will be based on the sensor cross-sensitivity, environmental response, calibration, and drift and bias estimation.

Thus the data management architecture requires capture and curation of original readings, translated values, and (where necessary) corrected values—along with metadata describing these transformations. Beehive pushes these data into Plenario. Plenario's application programming interfaces (APIs) can be used to build portals, to stream data, to pull data in batches, or to develop applications and workflows.

6 PUBLIC ENGAGEMENT AND POLICIES

The AoT project aims to create a community asset, where residents and businesses see the instrument as a public resource. Although Chicago has thousands of public cameras that garner little attention, the AoT project aspires to go beyond "accepted" to being "embraced."

6.1 Public Engagement

As initially conceived, AoT did not include cameras, and the microphone was envisioned to provide only sound intensity, with no potential for analyzing audio content. Adding these in response to science community input also introduced the potential for surveillance, and therefore the appearance of surveillance irrespective of the operating policies. This expectation was underscored in 2014 when the project team began to engage Chicago residents. At that time, the project was considering counting unique Bluetooth addresses as an approximation for pedestrian count. This concept was characterized in local media as "collecting data from" personal devices [18], demonstrating the importance not only of transparency but also of clear communication with residents.

Draft governance and privacy policies were created in 2015 and reviewed by privacy, technology, and legal experts in early 2016. AoT then partnered with *Smart Chicago Collaborative*, a local organization whose mission is to help residents make use of technology, to organize public meetings and online interaction via the OpenGov Madison "policy co-creation platform." The result was a set of community-vetted policies and governance structures, a report [19] detailing community input, suggestions, questions, and responses from the AoT team, and ongoing engagement processes for AoT operation.

6.2 Governance and Privacy Policies

AoT is a partnership in which Argonne technology is being adapted, with funding from the University of Chicago (UChicago) and the National Science Foundation (NSF), for use in urban areas. NSF funding is from the "Major Research Instrumentation" program and also supports the manufacture of 500 nodes and the associated Beehive management and data infrastructure. The City of Chicago is providing labor for installation, power, and permission to place the devices on public infrastructure. UChicago owns the devices.

As with all NSF-funded projects, the AoT cooperative agreement between UChicago and NSF includes a long-term data management plan. This plan involves a commitment to keeping all of the data open and available at no cost. Although the City of Chicago would typically request ownership of data collected using public infrastructure, the City agreed to grant UChicago ownership of the data given the alignment of the AoT cooperative agreement with the City of Chicago's open data policy (including not monetizing public data). All policies are available on the Internet at ArrayOfThings.org.

6.2.1 Governance Policy. AoT governance structure is implemented in three bodies: (a) Executive Oversight Council (EOC), (b) Scientific Review Group (SRG), and (c) Technical Security and Privacy Group (TSPG). The AoT project team will regularly update these groups regarding progress and new challenges or opportunities.

The *EOC* is co-chaired by the AoT project director and Chicago's Commissioner for Innovation and Technology, and includes individuals with varying perspectives and backgrounds, including academia, industry, policymakers, and community organizations. The EOC will provide guidance to the project regarding policy and public engagement, along with an approval process for policies and major operational changes such as related to privacy, data access, or installation and location selection strategies.

The *SRC* provides feedback and guidance to the project team and the EOC regarding the scientific and technical directions and services provided by AoT. The SRC will provide evaluation of major technical changes in terms of their merit in enhancing the scientific utility of the instrument. The *TSPG* role is described in detail below.

6.2.2 Privacy Policy. As the majority of data from AoT is environmental, data is published openly. Data such as weather or air quality have no personal privacy implications. A privacy policy is, however, central to the use of cameras and microphones. The AoT privacy strategy involves three components: technical architecture and operation, transparency, and accountability.

AoT's technical architecture and operation protects the privacy related to images and sound by defining two modes of operation: *common* and *sampling*. In *common* mode, edge computing will perform pre-approved operations on images in-situ and then delete. These operations will typically involve the use of machine learning to extract features from images, for instance determining the number of vehicles or pedestrians flowing through an intersection or the percentage of the roadway that is covered with standing water.

Periodically nodes will be placed in *sampling* mode, where some images and sounds will be collected and transmitted to a secure server. Although such data is not considered sensitive personally

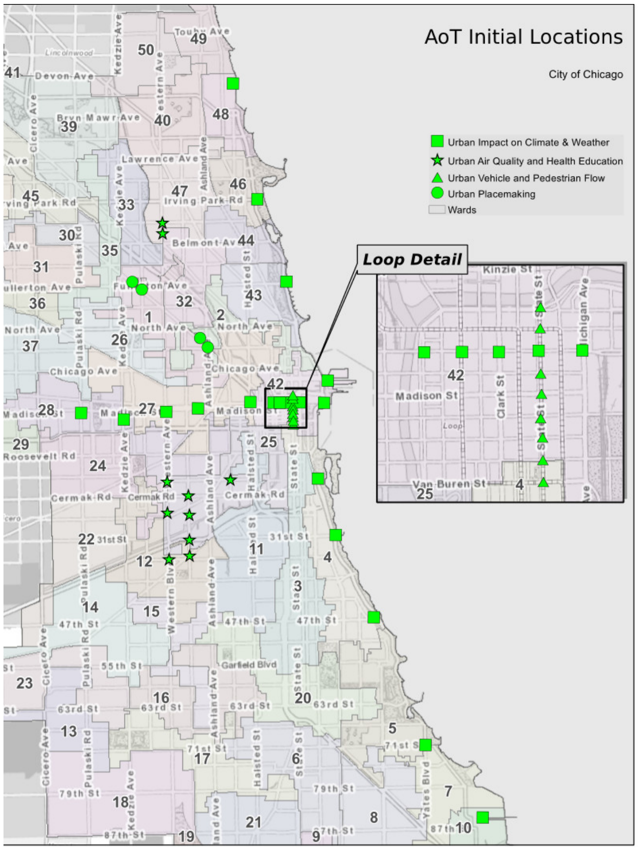

Figure 3: Subset of initial 40 AoT node locations. (City of Chicago).

identifiable information [13], it would be difficult to build trust with residents if images that might contain faces or license plate numbers were not protected. Thus access to sampled images and sound will be controlled, with access only granted in context of a confidentiality agreement. Whether or not this policy is strictly necessary with current systems, technology upgrades such as higher resolution cameras will eventually increase the sensitivity of images considerably [26] and the policies anticipate this potential.

Transparency was discussed earlier in context of public engagement, but also includes the availability of the Waggle software as open source [3]. Further, the selection of sites for AoT nodes is driven by a rubric that prioritizes for (a) interest or concern on the part of residents, (b) an engaged science team interested in analyzing the data, and (c) a city official (alderman, department head, commissioner) with interest in the analysis. Fig. 3 provides example sites with the first units.

Finally, critical to privacy is accountability. The TSPG operates independently of the AoT program operators. In consultation with NSF, we requested assistance from the Center for Applied Cybersecurity Research at Indiana University to convene and operate the TSPG.

TPSG will review AoT privacy policies and any proposed changes, advising the EOC on their findings. TPSG also has the ability to

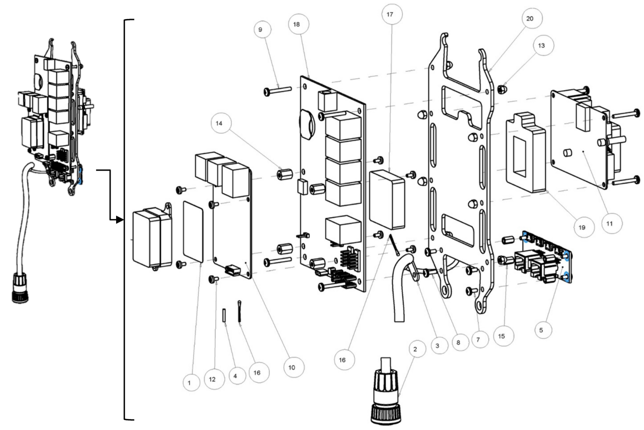

Figure 4: WagMan and Linux SBCs subassembly. (David Carhart, Product Development Technologies LLC)

audit the project, for instance requesting access to servers or nodes to validate that their software matches the published code at the GitHub site [7].

7 INITIAL MANUFACTURING AND DEPLOYMENT RESULTS

The most common sensor network architectures involve a sensor connected to a very reliable microprocessor, which reads the sensor and uses a modem to send the readings to a central server. Sensor nodes of this type can be integrated onto a single circuit board and inserted into a simple enclosure. Such architectures only support one of the three AoT use modalities (measurements). The addition of a programmable SBC introduced a host of reliability issues, as addressed in § 3.5, necessitating a supervisor board that could compensate for SBC hardware and software failures, facilitate updates, etc. But these features also increased complexity in packaging and assembly.

The transition from in-house assembly (by developers) to third party assembly revealed a number of areas in which this increased complexity affects assembly (and thus costs). First, WagMan has multiple connections (heartbeat monitor, temperature sensor, boot select) to each Linux SBC, and each of these represent a potential assembly error.

Second, for aesthetic reasons we minimized the size of enclosure, making internal assembly (Fig. 4) a “3d puzzle” due to connector clearances and heat management. For the latter, internal enclosure volume is insufficient for cooling via standard passive (fins) or active (fans) heat sinks. Thus WagMan is sandwiched between the two SBCs, which have solid aluminum block heat sinks. The blocks are precisely sized to contact the aluminum enclosure using foam padding to provide slight pressure. Because SBCs are not protected by an enclosure, their heat sinks are attached with adhesive that prevents their removal (without damaging the CPU chip). This necessitated special Odroid orders with no heat sinks. Further, a 2mm miscalculation in the aluminum block dimensions in early units resulted in blocks that were too deep. Assembled nodes passed all tests until they were screwed shut, at which point pressure from the extra 2mm of heat sink depth damaged the CPUs. In some cases this caused immediate failure, but in other cases the damage was not evident until the units had undergone several changes in temperature after being installed. This revealed the importance of rigorous testing, including temperature cycling, prior to installation.

The hand-off to third party assemblers also meant refactoring all of the quality control tests and instructions, which were originally written for project members who understood the systems. This meant that all test error indicators had to indicate what action was necessary, if any, to correct the error. For instance, some errors are informational and need not stop the assembly process (e.g., a software version error), while others require a halt in assembly (e.g., a component failure). In general, the hand-off to third party assemblers also drove the development of a prescribed assembly sequence to minimize the impact of subcomponent failures. To this end the assembly involves separate subassembly and unit test for the "brains" (4) and for the sensor harness, followed by testing these together, and finally testing after full assembly.

8 CONCLUSIONS AND NEXT STEPS

We initially planned to assemble the first 100 units in-house. However, in order to satisfy the demand from developers and for testing in other cities, we started the transition to third party assembly after only assembling a handful of units in-house. This change introduced six-month delay in our installation schedule, but also revealed critical design issues that might otherwise caused significant reliability issues with dozens of installations. These insights can now be translated directly into the design update, including some of the streamlining that will make assembly less complex, less costly, and more reliable. The initial assembly partner is under contract for 75 units, and in the next Request for Quotations for the selection of a higher-volume assembly partner we will include a design update to this end.

Our long-term vision for the project is to continue to refine the strategy of turnkey systems, where partners outside of Chicago can leverage the management and data pipeline in Chicago and the Plenario data dissemination platform. This has the added advantage that any applications developed with the Plenario API (§ 3.6) will work for residents of Seattle, Denver, or Portland (the first cities to deploy units) as well as for those in Chicago. This vision goes beyond AoT and Waggle hardware in that we have approached other sensor network projects, as well as companies, to discuss importing their data into Plenario. Similarly, we have worked with companies including BigBelly and Sidewalk Labs to enable their units to host the MetSense, LightSense, and ChemSense boards, further expanding the availability of urban data beyond the footprint of AoT and Waggle.

Finally, although the project has documented all of the sensors being used, including specific part numbers from which data sheets can be obtained, a more detailed set of documentation is in preparation regarding accuracy, precision, and calibration testing of sensors within AoT units.

ACKNOWLEDGMENTS

The AoT project is funded by an NSF Major Research Instrumentation grant (NSF 1532133 [8]) with in-kind cost sharing from the City of Chicago, Intel, AT&T, Cisco, Microsoft, Motorola Solutions, Schneider Electric, and Zebra. Waggle [4] was developed with funding from Argonne and UChicago. Plenario was developed with funding from NSF, UChicago, and the MacArthur Foundation. AoT would not have been possible without consistent support from Chicago Mayor Rahm Emanuel, Deputy Mayor Steve Koch, CIO Brenna Berman and Commissioner of Transportation Rebekah Scheinfeld; and Argonne Associate Laboratory Director Rick Stevens.

AoT co-investigators also include K. Cagney (UChicago), M. Papka (Argonne/NIU), D. Work (UIUC). The Waggle/AoT team includes C. Bailey, J. Bracho, W. Catino, W. Engler, V. Ghate, W. Gerlach, R. Kotamarthi, P. Lane, S. Shahkarami, and multiple summer student cohorts (http://wa8.gl), Additional partners include A. Emmanuel, K. O'Malley, M. Buchanan, and R. Ramos (Chicago DOT), P. Lukose and V. Patel (Surya), D. Pancoast, R. Drinkwater, and S.M. Basu (SAIC), D. Carhart (PDT, Inc.), and M. Potosnak (DePaul).

REFERENCES

[1] Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, and Michael Isard. Tensorflow: A System For Large-Scale Machine Learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI). Savannah, Georgia, USA*.

[2] Algirdas Avizienis. The Hundred Year Spacecraft. In *Evolvable Hardware, 1999. Proceedings of the First NASA/DoD Workshop on*. IEEE, 233–239.

[3] Pete Beckman, Rajesh Sankaran, and Charles E Catlett. 2014. Waggle: An Open Sensor Platform For Edge Computing. (2014). https://github.com/waggle-sensor/waggle

[4] Pete Beckman, Rajesh Sankaran, Charles E Catlett, Nicola Ferrier, Robert Jacob, and Michael Papka. Waggle: An Open Sensor Platform For Edge Computing. In *SENSORS, 2016 IEEE*. IEEE, 1–3.

[5] Justin H Breaux. 2015. New Sensor Array Changes The Data Collection Game. *Phys.org* (2015). https://phys.org/news/2015-03-sensor-array-game.html

[6] Michael T Carter, Joseph R Stetter, Melvin W Findlay, and Vinay Patel. 2013. Printed Amperometric Gas Sensors. *ECS Transactions* 50, 12 (2013), 211–220.

[7] Charles E Catlett, Pete Beckman, and Rajesh Sankaran. 2015. Array of Things. (2015). https://arrayofthings.github.io/

[8] Charles E Catlett, Peter H Beckman, Kathleen A Cagney, Daniel B Work, and Michael Papka. 2015. MRI:Development of an Urban-Scale Instrument for Interdisciplinary Research. (2015).

[9] Charles E Catlett, Peter H Beckman, Rajesh Sankaran, John Tolva, Alistair Boxall, Usman Haque, Philip James, Jack Westwood, J Stockton, and Richard Milton. 2015. *Instrumenting The City*. Report. www.urbanccd.org/events

[10] Charles E Catlett, Kathleen A Cagney, R George, S T Lindau, and D Pancoast. 2017. Urban Sciences Research Coordination Network 2017. (2017). http://usrcn.uchicago.edu

[11] Charles E Catlett, Tanu Malik, Brett Goldstein, Jonathan Giuffrida, Yetong Shao, Alessandro Panella, Derek Eder, Eric van Zanten, Robert Mitchum, and Severin Thaler. 2014. Plenario: An Open Data Discovery and Exploration Platform for Urban Science. *IEEE Data Eng. Bull.* 37, 4 (2014), 27–42.

[12] City Chicago. 2017. Chicago Traffic Tracker 2017. (2017). http://webapps.cityofchicago.org/traffic/

[13] DHS. 2012. *Handbook for Safeguarding Sensitive Personally Identifiable Information*. Report. https://www.dhs.gov/sites/default/files/publications/privacy/Guidance/handbookforsafeguardingsensitivePII_march_2012_webversion.pdf

[14] Beth Drewniak, Rao Kotamarthi, Rob Jacob, F Chen, and Charles E Catlett. 2014. (2014). www.urbanccd.org/events

[15] Etimad Fadel, Vehbi C Gungor, Laila Nassef, Nadine Akkari, MG Abbas Malik, Suleiman Almasri, and Ian F Akyildiz. 2015. A Survey On Wireless Sensor Networks For Smart Grid. *Computer Communications* 71 (2015), 22–33.

[16] Marcus Hahnel and Hermann Hartig. Heterogeneity By The Numbers: A Study Of The Odroid Xu+ E Big. Little Platform. In *Proceedings of the 6th USENIX conference on Power-Aware Computing and Systems*. USENIX Association, 3–3.

[17] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional Architecture For Fast Feature Embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 675–678.

[18] Rajiv Lal and Scott Johnson. 2017. Chicago and the Array of Things: A Fitness Tracker for the City. *Harvard Business Review* (2017).

[19] Denise Linn and Glynis Startz. 2016. *Array of Things Civic Engagement Report.* Report. https://arrayofthings.github.io/engagement-report.html

[20] Krupali Mistry and Avneet Saluja. 2016. An Introduction to OpenCV using Python with Ubuntu. (2016).

[21] Igor Paprotny, Frederick Doering, Paul A Solomon, Richard M White, and Lara A Gundel. 2013. Microfabricated Air-Microfluidic Sensor For Personal Monitoring Of Airborne Particulate Matter: Design, Fabrication, And Experimental Results. *Sensors and Actuators A: Physical* 201 (2013), 506–516.

[22] Bushra Rashid and Mubashir Husain Rehmani. 2016. Applications Of Wireless Sensor Networks For Urban Areas: A Survey. *Journal of Network and Computer Applications* 60 (2016), 192–219.

[23] Karl Rupp. The OpenCL Library Ecosystem: Current Status and Future Perspectives. In *Proceedings of the 4th International Workshop on OpenCL*. ACM, 13.

[24] Robert J Sampson. 2012. *Great American City: Chicago and the Enduring Neighborhood Effect.* University of Chicago Press.

[25] Mahadev Satyanarayanan, Zhuo Chen, Kiryong Ha, Wenlu Hu, Wolfgang Richter, and Padmanabhan Pillai. Cloudlets: At The Leading Edge Of Mobile-Cloud Convergence. In *Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on.* IEEE, 1–9.

[26] Lisa Thalheim, Jan Krissler, and Peter-Michael Ziegler. 2002. Body Check. *translated by Robert W. Smith), http://www. heise. de/ct/english/02/11/114* (2002).

[27] Geoffrey West, Luis Bettencourt, and Jose Lobo. 2013. *How Far Can Big Data Take Us Towards Understanding Cities.* Report. Santa Fe Institute. https://www.santafe.edu/events/how-far-can-big-data-take-us-towards-understa

[28] William H Whyte. 2012. *City: Rediscovering the Center.* University of Pennsylvania Press.

[29] Thomas Winkler and Bernhard Rinner. Sensor-Level Security And Privacy Protection By Embedding Video Content Analysis. In *Digital Signal Processing (DSP), 2013 18th International Conference on.* IEEE, 1–6.

[30] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. U-Air: When Urban Air Quality Inference Meets Big Data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1436–1444.