Array of Things User Workshop

August 29-30, 2018
Argonne, Illinois

Program Committee

Kathleen Cagney[1], Charlie Catlett[1,2], Pete Beckman[2,3], Kate Kusiak Galvin[1], Michael Papka[2,4], Mark Potosnak[5], Daniel Work[6], Douglas Pancoast[7]

Reporting Team

Program Committee, William Barbour[6], Jennifer Dunn[3], Nicola Ferrier[1,2], Vince Forgione[1], Derek Gloudemans[6], Rao Kotamarthi[2], Robert Mitchum[1], Rajesh Sankaran[2,] Von Welch[8]

Abstract: Objectives and Overview

Array of Things is a cyberinfrastructure instrument supporting three general types of research into smart and connected communities: (a) high spatial-temporal resolution measurements of the urban environment (air quality, noise, etc.), (b) rapid deployment and test of new edge technologies (measurement, communication, computation) at urban scale, and (c) research and development enabled by edge programmable devices embedded throughout a city and integrated with high-performance cloud services.

This workshop provided an update the AoT community—science users, technology partners, urban stakeholders, and sponsors—on the progress and status of the AoT instrument and the underlying technology and highlighted science projects that are beginning to incorporate the instrument, or its underlying Waggle edge computing platform. Round table discussions provided feedback on the instrument and the data services, as well as conveying "lessons learned" regarding the installation of science cyberinfrastructure in the public way. Tutorials covered data analysis tools, edge programming features of the instrument, and an AoT-inspired high-school and middle-school curriculum.

Invitations were extended to the combined participant lists for all previous workshops as well as to the many individuals who had expressed interest in the AoT Chicago cyberinfrastructure instrument, the Waggle platform, or creating an AoT instrument. Over 100 scientists, technologists, policymakers, and other stakeholders participated in the workshop along with over 50 remote participants.

Key outcomes: (a) a high-level outline of existing and new science requiring the AoT cyberinfrastructure instrument; (b) user feedback on enhancements to the data services and instrument roadmap; and (c) "lessons learned" regarding partnerships with stakeholders—from local government to residents to youth—necessary to deploy *persistent, production cyberinfrastructure* in the public way.

[1] University of Chicago
[2] Argonne National Laboratory
[3] Northwestern University
[4] Northern Illinois University
[5] DePaul University
[6] Vanderbilt University
[7] School of the Art Institute of Chicago
[8] Indiana University

C. Catlett (cec@uchicago.edu)

Science Talks, Tutorials, and Abstracts

M. Berkelhammer (UIC), M. Berman (UChicago), M. Brown (UIC), K. Cagney (UChicago), D. Chattopadhyay (UIC), S. Derrible (UIC), N. Ferrier (UChicago/ANL), J. Jiang (UChicago), A. Johnson (UIC), M. Kolak (UChicago), R. Kotamarthi (ANL), D. Liebovitz (UChicago), J. Lin (UIC), R. Mansour (Chicago Dept. of Public Health), E.G. Marai (UIC), R. Muehleisen (ANL), C. Negri (ANL), D. Niyogi (Purdue), A. Packman (Northwestern), A. Padmanabhan (UIUC), H. Pan (UIUC), V. Rivera (Northwestern/ANL), P. Schmid (Purdue), G. Sever (ANL), A. Sharma (Notre Dame), C. Song (Purdue), H. Swift (ANL), S. Wang (UIUC), V. Welch (Indiana), D. Work (Vanderbilt), D. Wuebbles (UIUC), H. Zhang (UChicago), B. Zhao (UChicago)

Acknowledgements

Array of Things (AoT) is funded through a Major Research Instrumentation grant from the U.S. National Science Foundation (NSF #1532133; supported by NSF CISE, ENG, and OIA) [1].

AoT cost-sharing partners include the City of Chicago, the University of Chicago, AT&T, Cisco, Intel, Microsoft, Motorola Solutions, and Schneider Electric. AoT technology partners include JCDecaux, Sidewalk Labs, Astronics, and Surya Electronics. The Waggle platform was developed with funding from Argonne National Laboratory, with extensions and customizations funded by the University of Chicago and through the AoT NSF MRI grant.

The project has benefited significantly from leadership and participation in key strategic and governance teams.

Executive Oversight Committee: Brenna Berman (CityTech), Don DeLoach (Centri), Danielle DuMerer (City of Chicago), Glenn Eden (Weber Shandwick), Aaron Koch (Trust for Public Land), Lynn Osmond (Chicago Architecture Foundation), Steven Philpott (XtraMedium), Ari Scharg (Edelson), Elissa Tenny (School of the Art Institute of Chicago), and Karen Weigert (Chicago Council on Global Affairs).

Science Advisory Group: Dan Reed (University of Utah), Pete Beckman (ANL/NU) co-chairs;

Security and Privacy Team: Von Welch (Indiana University), co-chair.

The NSF Cybersecurity Center of Excellence (NSF 1547272) and the Indiana University Center for Applied Cybersecurity Research provided guidance and leadership in cybersecurity and privacy.

CityTech (formerly SmartChicago Collaborative) provided leadership and coordination for the public engagement strategies and activities.

Beyond cost sharing, the AoT project would not be possible without the extensive support from, and partnership with, Argonne's Director, Office of the Director, and Associate Director for Computing, Environment, and Life Sciences; the City of Chicago Department of Innovation and Technology, Department of Transportation, and City of Chicago Mayor and Office of the Mayor.

Photographs are courtesy of Argonne National Laboratory except as otherwise noted.

Contents

C. Catlett (cec@uchicago.edu)

<this page intentionally blank>

1 Update on the Array of Things Project

Array of Things is a cyberinfrastructure instrument supporting three general types of research into smart and connected communities: (a) high spatial-temporal resolution measurements of the urban environment (air quality, noise, etc.), (b) rapid deployment and test of new edge technologies (measurement, communication, computation) at urban scale, and (c) research and development enabled by edge programmable devices embedded throughout a city and integrated with high-performance cloud services.

A central purpose of the AoT User Workshop was to provide users and partners with an in-depth update on the project status and roadmap through plenary presentations on (a) the deployed instrument, (b) the underlying technology, (c) the data services, and (d) efforts related to validating the measurements. Below we provide the project background and a summary of each of the four plenary presentations. All presentation materials for these are available online [2].

1.1 Background

Beginning in August 2013, collaborators from the University of Chicago, Argonne National Laboratory, and other universities convened to discuss the potential for improving atmospheric modeling for urban environments by deploying hundreds of sensors. Over the next three years a series of such workshops [2] with different science communities led to the design of a multi-sensor, programmable device for urban measurement, named the “Array of Things” to reflect the use of “Internet of Things” concepts to build an equivalent of an “Array” telescope--an instrument that combines measurements from many identical instruments in order to provide higher spatial and temporal resolution. Scientists from several dozen universities and national laboratories participated in all aspects of the instrument conceptualization and design through these workshops and ad-hoc meetings and discussions.

Concurrently, Argonne National Laboratory developed the open source Waggle [3] hardware/software platform to support the deployment of cyberinfrastructure in remote and/or harsh conditions, supporting remote programmability (“edge computing”) for data processing in situ, for tasks ranging from data reduction to machine learning for computer vision. To this end, Waggle platform employs resilient hardware/software principles.

In September 2015 the first AoT community workshop [4] was held in Chicago, at the start of a 3-year Major Research Instrumentation grant from the U.S. National Science Foundation [1]. At that workshop some 90 participants from around the world provided input into the design plans (that were derived from previous science workshops) for Waggle-based AoT nodes that would be deployed in hundreds of locations throughout the City of Chicago. Community and partner input through the breakout discussions of that workshop influenced the course of the MRI project in multiple ways, including (a) a regional federation model for the Waggle central services, (b) a focus on data use modalities that prioritize longitudinal data bulk download as well as semi-real time (latency of minutes) access to data via application programming interfaces, and (c) a programming environment for edge computation that supports commonly used machine learning frameworks initially, with a long-term objective to support a container-based approach. During the subsequent three years, the AoT team deployed two generations of test systems, in each case doing three-season evaluations and addressing challenges ranging from moisture and condensation protection to mounting procedures to power and cellular reliability issues.

With increasing interest from other universities and cities, the AoT team also accelerated the transition of node assembly to commercial partners. In late 2017 an industry partner assumed full responsibility for node assembly, and in mid-2018 the same partner also incorporated inventory and supply chain. Consequently, the AoT/Waggle team now places purchase orders for new nodes, specifying modifications and component substitutions as necessary to continue to improve and augment the platform or to procure different configurations for other research projects, such as environmental, mobile, or specific industry configurations. Finally, in addition to the standard AoT node configuration and packaging, a version of the nodes can be ordered for developers, with enclosures that are readily opened and with additional space for expanding the electronics or sensor complement.

With over 100 nodes installed, data services initiated, and a second wave of 100 nodes to be installed in late 2018, the AoT team organized a second community workshop—the AoT User Workshop, hosted by Argonne National Laboratory outside of Chicago, Illinois.

1.2 Instrument Status and Design Principles

AoT project principal investigator Charlie Catlett presented the motivation and background of the project, its goals, and an update on the functions and deployment status. The AoT instrument is designed to (a) provide unprecedented spatial and temporal resolution measurements of the urban environment (weather, air quality, sound, etc.), (b) serve as a flexible test facility to evaluate new sensor technologies at urban scale, and (c) provide a cyberinfrastructure testbed instrument via remotely programmable "edge" computing capabilities for data processing in-situ, including research in capabilities such as "intelligent" infrastructure, and image and sound processing such as using machine learning techniques to "measure" flooding or the flow of pedestrians and vehicles.

After several successive deployments of dozens of test nodes, the instrument began to transition from design and testing to full-scale cyberinfrastructure in early-2018 with 105 nodes installed throughout Chicago. As of mid-2018, smaller test instruments (4-10 nodes) were being installed by partners in Seattle (University of Washington), Portland (City of Portland, funded through the NIST Global Cities Team Challenge program), Denver (Panasonic), Chapel Hill (University of North Carolina), and Syracuse (Syracuse University). Nodes outside of Chicago are funded through research agreements between partners and the University of Chicago. Node management and data aggregation are done in Chicago as part of the research agreement.

At the same time, the project began to publish data for bulk download as well as via application programming interfaces for portal and application developers. A multi-institutional air quality sensor evaluation team was formed to rigorously evaluate and characterize the experimental air quality sensors through collocation at regulatory (Federal Reference Method) instrument sites, beginning with a site on the South Side of Chicago, operated by the Illinois EPA and Cook County Department of Environmental Controls. (§1.5)

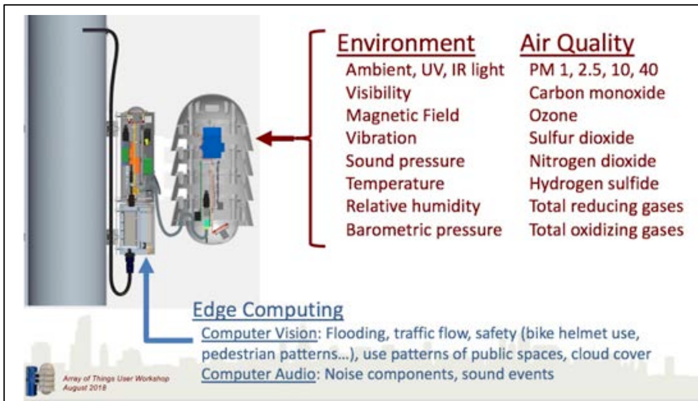

Figure 1: Cutaway view of current AoT node configuration showing sensors. Sensor pod is expandable to accommodate new experiments; deployment in groups of 100 supports rapid evolution of sensor complement and edge computing hardware.

AoT nodes are assembled by a commercial partner, Surya Electronics,

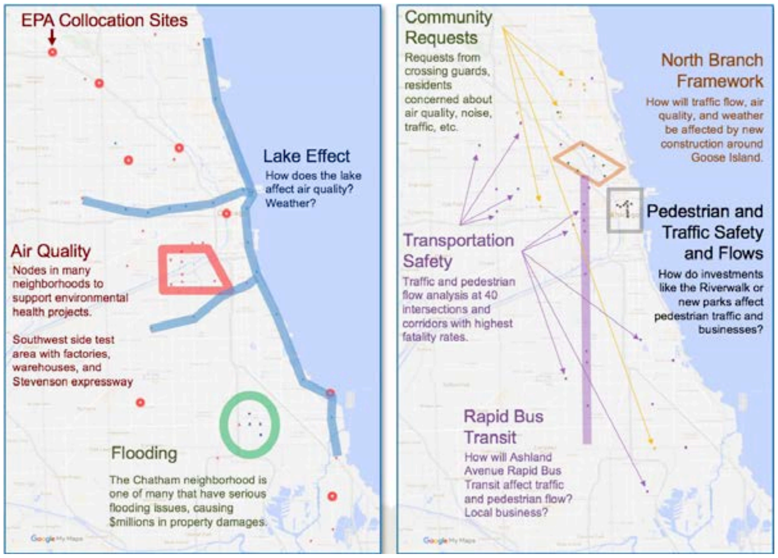

Figure 2: Initial 105 AoT node locations, showing that locations are selected in groups as part of specific science investigations.

in Glendale Heights, Illinois. A batch of 150 nodes will be assembled beginning in early September 2018, 100 of which will be installed between October and December 2018 in Chicago. The additional 50 nodes are being provided to partners on a first-come, first-served basis, with a number of research agreements in place or in progress, involving both cities in the U.S. (Denver, Nashville) and around the world, including Bristol (UK), Taichung City (Taiwan), Vancouver (Canada), Melbourne (Australia), and Hong Kong.

Data services include bulk download in various forms (all data as well as subsets by recent day, week, or month), a growing number of tools for extracting or subsampling these data packages, and APIs for more specific queries, including streaming access and queries for near-real time measurements across multiple nodes or by node.

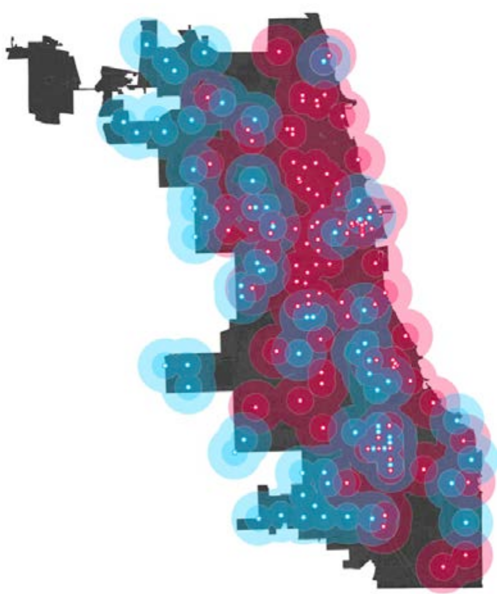

Figure 3: Current (red) and 60 of 100 additional planned (blue) AoT nodes. Both 1km and 2km buffers are shown, illustrating that even with 200 nodes over 95% of Chicago's residents will live within 2km of a node and over 75% will live within 1km. (GIS map created by A. Laha, Center for Spatial Data Science, University of Chicago)

Locations selected for the Chicago instrument are driven by specific science and/or policy questions and experiments, thus form a collection of discrete "experiments," each with some number of nodes (Figure 2). However, *all nodes have the same configuration* and thus an experiment aimed at traffic measurement will also provide data related to microclimate and air quality. With the currently deployed 105 nodes, over 80% of Chicago's population resides within 2km of a node and 42% of the population lives within 1km. With the next 100 nodes over 95% of the population will live within 2km and over 75% will live within 1km (Figure 3).

1.3 Open Waggle Technology Platform and Extensions

Argonne created the Waggle project in 2013 to develop a modular open source hardware/software platform for the deployment of instruments, with in-situ (edge) computing capability, in remote and/or harsh locations. The platform is designed to support scientific cyberinfrastructure, such as AoT, providing a configurable node design, simple "plug-in" scripts for adding and sampling sensors (or other devices, e.g.

radios or cameras), a reliable data pipeline for aggregating data in a central database, and secure management capabilities and tools for node management as well as for managing user-written software modules such as for edge-supported image analysis. Extensions to the node controller hardware in 2018 also accommodate additional edge computing hardware.

The Waggle project is led by AoT co-principal investigator Pete Beckman and emphasizes student participation through annual undergraduate and graduate student fellowships involving 10-15 students. Beginning in 2018 the Waggle summer program expanded to include high school students.

The Waggle platform aims to exploit the combination of (a) emerging low-cost, high-quality sensor technologies, (b) new computational capabilities in ever-smaller form factors, including hardware optimized for machine learning, and (c) hardware and software techniques for ensuring reliability and resilience of such technologies in remote and/or harsh installations. Waggle "nodes" operate autonomously with no open ports; thus, connectivity must be initiated by the nodes. Nodes connect only to a central "beehive" where they push data and receive instructions (such as to update software). Nodes report their measurements at regular intervals, typically set to 30 seconds.

Data is reported and stored in a simple CSV structure that includes timestamp, node ID, sensor parameter, the electronic sensor reading as well as a translation of that reading into intended units (e.g., degrees C or parts-per-billion). Metadata is used in concert with reference data embedded in the data downloads and at the data publication site, which includes tables such as those mapping node ID to physical location[9], an ontology to describe the sensor (or other device), including associated data sheets, model numbers and board firmware level. Tools for extracting and transforming the data are also provided at the bulk download site. (See §1.4)

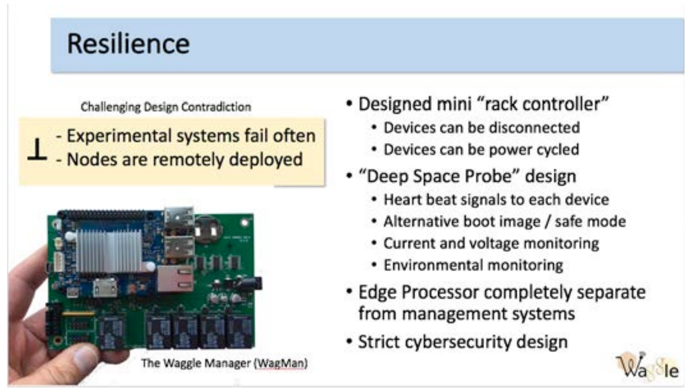

Figure 4: Waggle manager board (WagMan) implements a suite of resilience capabilities including the ability to fully rebuild the edge computing hardware (Linux systems) and to hibernate systems during high-heat or other hazardous conditions.

In addition to the Waggle platform, Argonne created a compatible platform for deploying more targeted sensing capabilities in greater numbers and at much lower cost: microWaggle. Based on commercial Particle.io wireless-connected microprocessors, the microWaggle units report data using the same CSV format and metadata conventions as standard Waggle nodes, thus their data can be seamlessly mixed with data from standard Waggle nodes such as in AoT. MicroWaggle units have been developed for moisture sensing as well as for mobile air quality measurement experiments (§2.1.10).

1.4 Data Services

All AoT data is open and free, owned and maintained as such by the University of Chicago. AoT data services target three general dissemination channels as part of a strategy to make data legible in context of diverse disciplines, analytics resources, and expertise. First is bulk data download, modeled after other longitudinal data sources such as for air quality or weather. Second is an application programming interface (API) suite that implements a powerful set of queries, intended to support

[9] Waggle supports GPS for mobile deployments, where node ID does not map to a fixed location.

application developers who wish to access semi-live (2-5 minutes latency) data. Third is a partnership program with groups who operate data analytics services, such as a community-specific science gateway or portal, or a group providing geographical information service. The API described below supports both applications and portal partners.

AoT data is captured in the individual nodes and transmitted at roughly 30 second intervals to the central "Beehive," which includes a Cassandra database. Data is published on a public bulk data download server every day at midnight UTC. To support API users and data streaming to portals or science gateways, Beehive also exports an m-minute window of data every n-minutes, with these values currently set to 30 and 5 (most recent 30 minutes of data pushed every 5 minutes).

1.4.1 Data Formats

Each bulk download file comprises a TAR archive with the following components:

data.csv.gz	Compressed file of all data values
nodes.csv	List of nodes in the dataset and their metadata
README.md	An explanation of the database fields
sensors.csv	A list of sensors and their metadata
provenance.csv	Metadata on the entire dataset archive
offsets.csv	data.csv.gz file byte offsets

The schema for these files is as follows:

CSV Data:	timestamp, node_id, subsystem, sensor, parameter, value_raw, value_hrf
Nodes CSV:	node_id, project_id, vsn, address, latitude, longitude, description, start_timestamp, end_timestamp
Sensors CSV:	ontology, subsystem, sensor, parameter, hrf_unit, hrf_minval, hrf_maxval, datasheet

These formats show that both "raw" values (the electrical reading from the sensor) and interpreted, or "hrf" (human readable format) values are included. The latter (hrf) represents the measurement as translated from the sensor reading (an electrical voltage or current level) and a translation function. In some cases, the translation is a linear interpolation, and in other cases it is a more complex translation that may include other sensor readings (e.g., temperature or humidity, if these affect the sensor's response) or values specific to the sensor, such as coefficients related to the specific sensor's properties (e.g., it's zero-value reading). In each case the sensor metadata and data sheets will include the information necessary to evaluate the translation from raw to hrf.

For several sensors, such as sound (reports sound pressure at 10 octaves) or vibration (using an accelerometer that provides x, y, and z forces) algorithms are being evaluated for translation into scientifically useful hrf values. Once these experiments are complete (late 2018) these formulae will also be provided as metadata. Note that additional or alternate algorithms can be added at later dates, at which time the raw values can be re-interpreted into new hrf values. As the instrument moves to production any such re-interpreted values will be published as updated versions of the original data, while previous data files will be retained as well.

1.4.2 Bulk Data Download

Beehive exports data to a public download site daily, enabling users to download either all data or select from a number of different subsets. Presently there are monthly files as well as the most recent five weekly and the most recent 8 daily files. At present, a basic set of tools is also provided given that these data files are large relative to a typical workstation (6-months of data from 100 nodes is of order 200 GB and 1.5B records). These enable users to export a range of dates or a subset of sensor

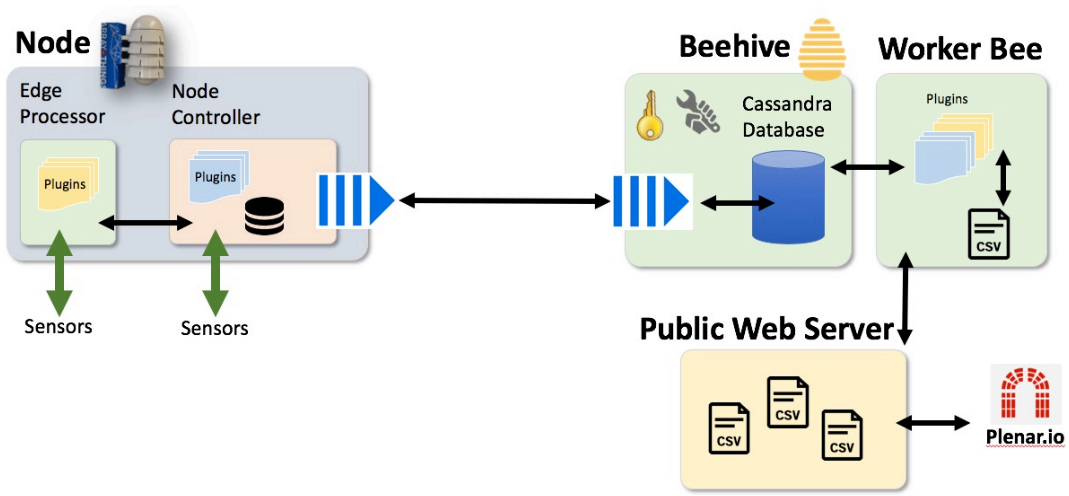

Figure 5: Waggle Data Pipeline showing AoT node, Beehive, and data services.

readings. Additionally, tools are provided to reduce the data by aggregating over time intervals (e.g. hourly, daily) using standard functions such as count, sum, mean, median, min, and max. Tools for computing moving averages as well as plotting data are also provided at the download site.

1.4.3 AoT Data API

Access to live (2-5 minute latency) data through the API is supported through the Plenario [5] open data discovery and exploration resource, originally developed through an NSF CISE grant [6] and extended with funding from the MacArthur Foundation and the City of Chicago (via a grant from Bloomberg Philanthropies). Plenario holds all AoT data, updated at 5-minute intervals. An initial API was tested during 2016-17 and updated in 2018 based on user feedback. The API [7] supports a range of filtering options, including:

- By project (AoT Chicago, partner AoT instruments in other cities, and projects using Waggle and/or microWaggle devices)
- By node or subset of nodes
- Time period
- Geographic location (a latitude-longitude or within an area defined by a polygon, such as generated by a map-based interface)

Data is returned in JSON or GeoJSON format. The API also supports streaming data (or subsets through the above filters) using the socket.io protocol. Client libraries for Python, R, and JavaScript are under development, to be made available through common language package managers (e.g., pip).

1.5 Experimental Sensor Evaluation Process

Each AoT node hosts a Chemsense board, a unique set of sensors that measure seven gases relevant to air quality. Chemsense was designed by Intel as part of their MRI cost-sharing partnership, and uses electrochemical gas sensors developed by SPEC Sensors, CA.

The chemical sensors are calibrated by the manufacturer and shipped for installation in the AoT module. To test the performance of these sensors after placing them in an AoT node, a collocation experiment was designed. We are currently conducting this study to evaluate the performance of these low-cost gas sensors with reference grade systems operated by the EPA. The collocation site is operated by Cook County on behalf of the US EPA for air quality standards conformance in the county. There are three gases measured by Chemsense that are also monitored by the EPA instruments: ozone, NO_2 and SO_2. The SO_2 levels measured at the site are relatively low and often below the detection limit of the Chemsense SO_2 sensor. As a result, we have focused our attention on understanding ozone and NO_2. There is considerable cross sensitivity between the ozone and NO_2 sensors. Three types of error were considered: random, systematic and spikes for evaluation with the reference instruments. For NO_2 and ozone, the random error is larger than the reference sensor, but the results are encouraging ($r^2 > 0.65$). There was some systematic error in the zero (residual from Chemsense compared to EPA data) that might be related to either sensor drift or temperature, but the level was relatively low and should be amenable to analysis with spatial techniques across the Array. We are continuing the analysis of spikes in the NO_2 signal that were not detected by the EPA reference sensor and were not related to temperature or absolute humidity. Overall, the results are very encouraging and demonstrate that the Chemsense boards produce data that can be applied to city-wide urban air quality science questions.

2 Science Highlights

Lightning talks from a diverse set of AoT science partners highlight the diversity of research enabled by the AoT cyberinfrastructure as well as leveraging the AoT instrument as a testbed for extensions of the underlying Waggle platform. Below are abstracts from these science lightning talks. In most cases the presentation materials are available online and linked to the agenda appended to this report. One presentation contained information not yet approved for public dissemination by the funding sponsor, and thus at the time of this report only the abstract is available for publication.

In addition to abstracts from lightning talks on the workshop agenda, many participants submitted abstracts detailing science that they are interested in pursuing with AoT.

2.1 Workshop Lightning Talks

2.1.1 Urban Microscale Modeling and Developing Heat Mitigation Strategies (Ashish Sharma, University of Notre Dame)

Extreme high temperatures in summers lead to an increased risk of heat-stress in vulnerable neighborhoods of cities. With intense socioeconomic, health and meteorological data collection in urban areas, there is a need to better understand the linkages between climate change and planned adaptation and mitigation efforts that impact urban metabolism at micro- (neighborhood) scales. Current urban heat maps cannot capture spatial trends and hot spots at fine spatial scales. This study analyzes in situ environmental and air quality observations in the City of Chicago using the newly installed, networked urban sensor project, AoT. Measurements from AoT nodes are spatially interpolated on a uniform grid using kriging techniques and is being used train simplistic models related to outdoor thermal comfort, energy consumption, climate change mitigation etc. We also use the City of Chicago's public dataset to obtain micro-scale social metrics, such as frequency of

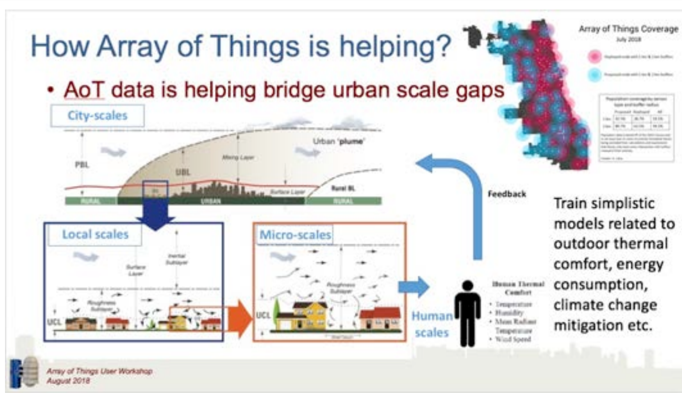

Figure 6: Urban Microclimate (Ashish Sharma, Notre Dame).

311 calls and crime, along with electricity consumption to develop a framework for urban planners to reduce adverse heat impacts. Long-term extrapolations will be addressed, along with implications of implementation of green infrastructure to combat urban heating in Chicago's most vulnerable areas. This work is funded by the Environmental Change Initiative, University of Notre Dame.

2.1.2 Understanding Environmental Effects on Behavior (Marc Berman, University of Chicago, Director, Environmental Neuroscience Laboratory)

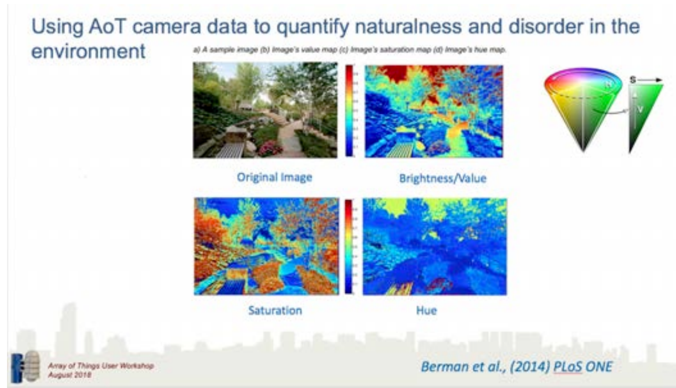

Figure 7: Behavioral Sciences (Marc Berman, UChicago).

The UChicago Environmental Neuroscience Laboratory studies the interaction of individual person factors (e.g., brain activity, psychological state, genomic variables) with external environmental factors (e.g., urban density, greenspace, disorder) and how those factors interact to influence brain activity and behavior. This lightening talk provided a series of studies showing how brief interactions with natural environments (e.g., a short walk in the park) can improve cognitive performance relative to brief exposures to urban environments (e.g., a short walk in an urban setting). We replicate these effects when people view pictures of nature and urban scenes and also hear nature and urban sounds. Importantly, these cognitive effects are not driven by mood, as even when participants do not enjoy the nature interaction, they still obtain the cognitive benefits. Results were also presented showing how neighborhood greenspace is related to health, controlling for age, education and income. These effects were most robust for public greenspace, i.e., trees on public easements separating street from sidewalk, compared to trees on private land (i.e., trees in people's backyards), presumably because more people in the neighborhood have access to those trees. These results have led to the development of an app, *ReTUNE* (Restoring Through Urban Natural Experiences), where we can provide individuals with the most restorative walking route between two points in a city that maximizes greenspace, while minimizing crime and sound. AoT data will enable us to extend our research beyond the laboratory with high ecological validity. We plan to use the camera and microphone on the AoT nodes to quantify naturalness, disorder and aesthetic beauty of different neighborhoods and how those variables may affect behavior such as walking speed, social interaction and crime. In addition, because of the dynamic nature of the data we can look at changes over time of day and season. We also plan to use the data to refine our app. Lastly, we hope that with our research we can inform the placement of new AoT nodes and also conduct intervention studies where we can manipulate greenspace in the environment and monitor changes in behavior in those environments. The work funded by NSF, the TKF Foundation, the John Templeton Foundation and the University of Chicago.

2.1.3 Computer Science, Transportation, and Environmental Research (Nicola Ferrier, University of Chicago and Argonne National Laboratory)

This presentation described some of the initial efforts to develop computer vision and machine learning applications for the AoT platform. Initial urban and environmental applications include transportation measures (traffic flow, pedestrian counting, aircraft recognition), detection of urban flooding, and monitoring plant growth on green roofs. The opportunities for basic machine learning research with the AoT instrument also include the potential for partnerships between computer scientists and domain scientists whose research requires analysis of imagery (as noted in §2.1.2 and

§2.1.7), sound, or of multiple sensors (e.g., adding vibration) to understand particular urban or environmental phenomena. Irrespective of the edge computing or storage resources of a given AoT node, supporting a suite of "measurements" in this fashion, such as counts of pedestrian or vehicles, or quantifying flooding, there will be computer science challenges related to areas ranging from scheduling to resource management, including the management of multiple user-developed software modules within the nodes. Simply put, the AoT instrument will not only support research opportunities in measurement and computer vision, but in fundamental computer science from operating systems to distributed machine learning architectures.

Figure 8: Computer Science (Nicola Ferrier, ANL/UChicago).

2.1.4 Neighborhood-Resolution Environmental Data for Aging Studies (Kathleen Cagney, University of Chicago)

The older adult population is rapidly growing, but little is known about how and to what extent older persons maneuver in their communities. In this NIH (NIA) and OBSSR-funded study we track older adults' movements—in, out of, and across their communities—to provide insight into their span of engagement, the contexts most relevant for their health and well-being, and their access to social and community resources. These dynamic data, enhanced with AoT data on the micro-environment, provide an unprecedented opportunity to examine social context effects across time – this allows researchers to situate residents in their differing social, physical and service environments and assess the unique impact of these spaces on health.

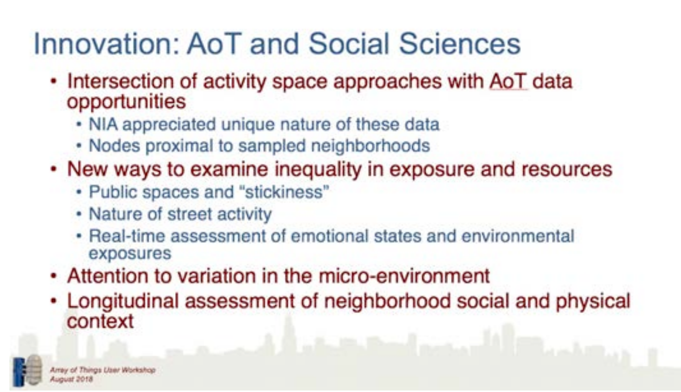

Figure 9: Social Sciences (Kathleen Cagney, UChicago).

2.1.5 AoT and Waggle to Support Atmospheric Sciences (Cristina Negri, University of Chicago and Argonne National Laboratory)

A potential use of AoT would be to provide dense sampling of an environmental system, in conjunction with sparse reference grade measurements of complimentary phenom logical variables, to reveal process mechanisms that are heterogeneous in space and time. Thus, an aspirational application of AoT and Waggle in the environmental sciences is its potential use for atmospheric research that is influenced by variability in the driving forces. We envision the potential for using AoT and Waggle nodes to understand aboveground and belowground processes that control the cycling of water at the Southern Great Plains, Oklahoma site of the Atmospheric Radiation Measurement (ARM) Research Facility. The water cycle in atmosphere is coupled to moisture transfer processes through the biosphere and soils, the nature and extent of the coupling giving rise to short-term and log-term droughts, some cases leading to sustained precipitation and floods. A tested, validated, and tailored set of Waggle nodes may allow researchers to better understand, for example, heterogeneities in the landscape at different scales and how these drive the atmospheric water cycle. The ARM User Facility is owned by the US Department of Energy Office of Science's Biological and Environmental Research, and managed by a partnership of nine National Laboratories, including Argonne.

2.1.6 Using AoT for Transportation Research (Dan Work, Vanderbilt University)

Motivated by an influx of residents, finite infrastructure capacity, and the rise of a new class of technology-driven transportation services, understanding urban mobility is now a critical concern of major cities worldwide. In this talk I will explore some examples of how the performance of transportation systems is measured today, and how this is likely to change in the near future due to advances in fixed and mobile sensing. The recent development and deployment of inexpensive and connected sensor nodes like the AoT provides a new perspective on mobility across modes while also allowing the impacts of mobility (e.g., air quality) to be quantified. This research is supported by funding from NSF Cyber Physical Systems; NSF Civil Infrastructure Systems; NSF Infrastructure Management and Extreme Events; NSF Partnerships for International Research and Education, NSF Major Research Instrumentation; the Siebel Energy Institute; and the US Department of Transportation Region 5 University Transportation Center.

Figure 10: Transportation Research (Daniel Work, Vanderbilt).

2.1.7 Integrating Environmental and Socioeconomic Data with Electronic Medical Records and Molecular Phenotypes (David Liebovitz, University of Chicago)

The impact of the environment on health and risks for chronic disease is well established for many conditions. Availability of electronic medical record data, clinical research databases such as from the COMPASS study [8], and insights provided by epigenetic analyses provide additional dimensions for analyzing the effects of micro-environmental exposures and health. Leveraging these data sources and working together with the AoT community will therefore provide opportunities to answer questions such as:

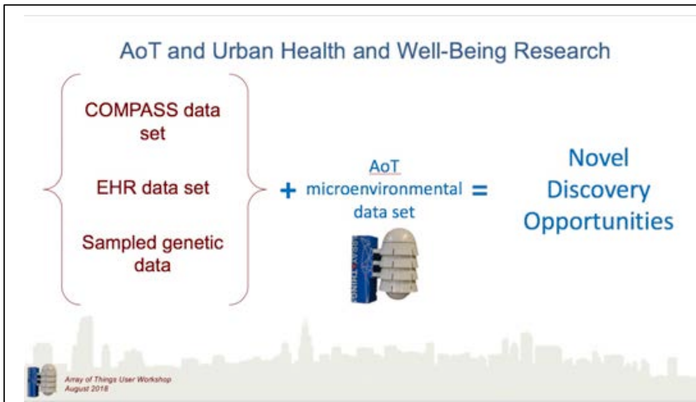

Figure 11: Urban Well Being (David Liebovitz, UChicago).

1. Whether **improved models for predicting morbidity and mortality** (e.g., for respiratory and cardiovascular morbidity, pregnancy complications, diabetes and hypertension control) are possible leveraging **event detection through ambulatory and ED data from medical records systems.**

2. Whether micro-environmental exposures contribute to **differential advances** of the epigenetic clock for sampled patients.
3. Determining the clinical impact of enriching EHR data with home address specific micro-environmental data for use **at the point of care**.

COMPASS is funded by the University of Chicago Medicine Comprehensive Cancer Center, one of two NCI-designated Comprehensive Cancer Centers in Illinois.

2.1.8 Partnership for Healthy Cities (Raed Mansour, Chicago Department of Public Health and Marynia Kolak, University of Chicago)

The Chicago Department of Public Health, with partners at the University of Chicago Center for Spatial Data Science and City Tech Collaborative, are collecting publicly available environmentally related data sets from a variety of federal, state, and local sources like the Chicago Data Portal, National Aeronautics and Space Administration's Open Data Portal, US Environmental Protection Agency Air Quality System, and the National Oceanic and Atmospheric Administration's Climate Data Online, amongst several others. The collected open data sets will eventually visualize baseline air quality on the Chicago Health Atlas which displays health-related information visually with maps, charts, and graphics so residents and experts can better understand health in Chicago. Through the use of mixed sensor modeling, the project aims to predict air quality in each grid cell (i.e., 1km) using NASA satellite sensors (MODIS/MAIAC estimate of Aerosol Optical Depth), EPA and the AoT sensors, as well as various other predictor variables like point emissions, area emissions, meteorological data, planetary boundary layer, green spaces, land cover, etc. This project is funded by The Partnership for Healthy Cities Grant led by Bloomberg Philanthropies and Vital Strategies.

2.1.9 Distributed high-frequency urban sensing for flood prediction and impact assessment (Aaron Packman, Northwestern University)

The combination of increasing urban population and changing climate increases water vulnerability for cities. In particular, extreme weather events such as floods and droughts represent major hazards for cities. Green infrastructure, including nature preserves, parks, greenways, and urban farms, can potentially provide low-energy solutions to flooding and water storage while also providing additional ecosystem services such as biodiversity, temperature modulation, recreation, food, and improved health. However, current green infrastructure designs are based on empirical approaches developed simply to enable greenspace to be constructed within urban environments, and have not been rigorously evaluated in terms of engineering performance.

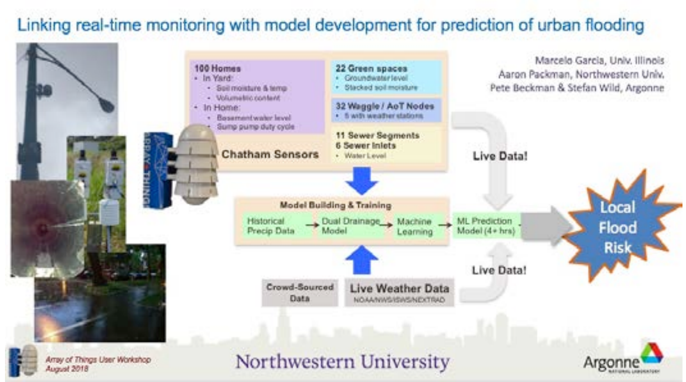

Figure 12: Urban Hydrology (Aaron Packman, Northwestern).

Using the AoT network as a backbone for sensing, data transmission, and archiving, we are observing the performance of green infrastructure for storm water storage, along with conditions that are important to maintaining human health and functional greenspaces in the urban environment. We have instrumented a network of nature preserves, green roofs, farms, and parks in Chicago. Waggle-based greenspace sensing includes atmospheric conditions, precipitation, surface and groundwater levels, and soil moisture. Results to date show tremendous variability in soil moisture and groundwater levels in urban greenspaces, yielding substantial storage of storm water and high variability in hydrologic conditions for plant growth. We are currently implementing additional measurement capability, including image-based evaluation of flooding in streets and parks, and instrumentation in and around buildings using new “micro-waggle” sensors (§2.1.10). Using this suite of methods, we have designed new neighborhood-scale sampling networks that can be used to diagnose causes of flooding in vulnerable Chicago communities. Our ongoing data collection, combined with simulation/prediction schemes, will enable green infrastructure to be designed to reduce flooding and other extreme weather hazards while providing strong co-benefits for biodiversity, recreation, and health.

2.1.10 microWaggle for Hydrology (Vivien Rivera, Northwestern University)

To support targeted sensor deployment such as moisture sensors in building basements or green infrastructure installations, the Waggle team developed microWaggle. MicroWaggle uses a wireless-enabled (WiFi, cellular, or mesh) commercial microcontroller from Particle.io to host such sensors. The microWaggle nodes report their measurements using the same schema as AoT nodes (§1.4.1), with timestamp, node_id, and other metadata and can be quickly tested with new experimental sensors in a manner similar to an AoT node. A relay server pulls data from the Particle cloud into Beehive, thus microWaggle measurements are seamlessly integrated with other Waggle deployments, including AoT, and can be analyzed using the tools and methods developed for other AoT/Waggle data. MicroWaggle nodes are remotely configurable, either individually or simultaneously in groups, enabling adaptive sampling or adjustment of measurement routines in response to new information in the sensor network.

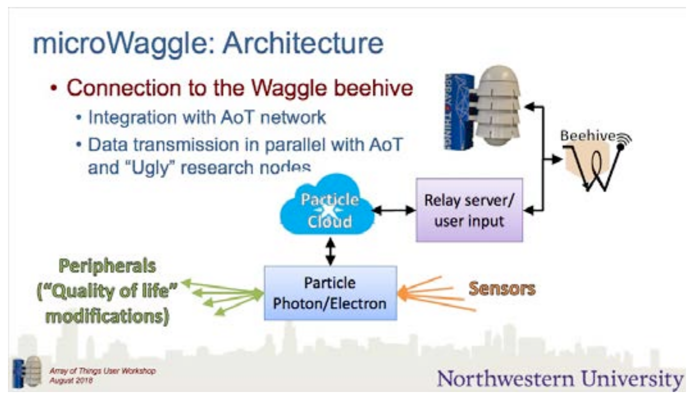

Figure 13: Sensor Network Architectures (Vivien Rivera, Northwestern/ANL).

To support flood modeling and prediction in Chicago’s Chatham neighborhood, microWaggle nodes have been equipped with soil moisture and water level sensors in residential and public building basements, and report using cellular data. This work is supported by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education (ORISE).

2.2 Additional Science Abstracts Submitted by Workshop Participants

With participation from over 60 users and potential users of the instrument, including many computer scientists with interest in exploiting and modifying the underlying platform, there was not time to accommodate lightning talks from each participant. In addition to lightning talks, many participants contributed abstracts outlining science uniquely enabled by the AoT instrument.

2.2.1 CyberSecurity Research for Smart and Connected Community Infrastructure (Von Welch, Indiana University)

The ResearchSOC (NSF award #1840034 / researchsoc.iu.edu) is an NSF/OAC-funded center which will deliver operational security services to NSF and other research projects. One of the challenges the ResearchSOC will tackle is unique scientific instruments, with the AoT a prime example of such instruments. As expressed in the letter of collaboration from ResearchSOC PI Welch, the AoT staff will work with the ResearchSOC staff to understand the challenges in adapting cybersecurity services for the AoT, such that the ResearchSOC is poised to implement appropriate support into its service offerings.

2.2.2 Using AoT's Waggle Platform for Edge Computing and Machine Learning (Heather Zheng, Ben Zhau, University of Chicago)

We are interested in design and experiment with outdoor event recognition systems based on AoT. It is well known that (sudden) event recognition can provide convenience, safety, and efficiency for human users. It allows an immediate response during emergency situations such as a fire, medical emergency, physical crimes, and abnormal human activities. AoT offers a great opportunity to develop and deploy low-cost city-wide event recognition using commodity, low-cost hardware. In particular, we are interested in using AoT nodes to obtain extra dimensions of environmental observations: images, sound, vibration, and radio spectrum usage, and processing these data directly on AoT nodes locally to enable real-time operation (thus fast reaction) and to preserve user privacy. These requirements, combined with the scaling factor (citywide), lead to significant open research challenges. Our work is funded by NSF, DARPA, and Google (faculty award).

2.2.3 High-Dimensional Information Visualization and Augmented Reality Applications for Urban Data (Maxine Brown, Andy Johnson, University of Illinois-Chicago)

The UIC Electronic Visualization Laboratory (EVL) has NSF MRI funding to build the Continuum instrument, an immersive visual analytics space that is instrumented with a variety of sensors (cameras, microphones, RFID, depth cameras, temperature, etc.) to make it easier for the room to help people interact with their data by knowing more about what is going on inside the workspace. To give context to what is going on inside the room, it would be helpful to know what's going on outside the room, at multiple distances, by getting information from AoT as well as other data sources (vehicle traffic, pedestrian traffic, light levels, weather, atmospheric conditions, news, etc.). Continuum is funded by NSF award #1625941, PI, Andrew Johnson, Co-PI Maxine Brown.

EVL also specializes in visual analytics and collaborative visualization and virtual-reality technologies. EVL is interested in visualizing AoT data over time and integrating it with other — slower to update — City of Chicago and Federal census data, for viewing on a variety of devices, from hand-held devices (e.g., smartphones and tablets) to large-scale tiled display walls (in order to see multiple representations of the data at the neighborhood and City level). In addition, EVL is interested in developing augmented-reality applications for hand-held devices (smartphones and tablets) and headsets that enable people to look at specific, individual AoT devices and view real-time data as it is collected. Proposed by Andrew Johnson, EVL Director of Research.

2.2.4 AoT to Explore IoT Distributed Systems and Resource Management (Junchen Jiang, University of Chicago)

The AoT instrument presents an interesting testbed for exploration of efficient and scalable distributed systems for deep learning analytics, with a focus on video data analytics. Additionally, the instrument is a potential real-world use case of geo-distributed data analytics and machine learning. A key challenge facing many IoT systems is reducing the cost of data communication between devices and the cloud, which often relies on expensive cellular data plans. The AoT project provides an excellent use case to test new approaches, in which different data sources require different levels of granularity, fidelity, and freshness. Updating data analysis models could potentially be another contributor to the communication overhead as the AoT platform begins to support more dynamic approaches to loading new models. Such flexibility will be important to support increasing numbers of applications and analytics tasks, creating new opportunities to amortize cost of data processing/storage across applications. Exploring these topics requires a flexible resource management layer, driving the mechanisms used in the Waggle platform used by AoT.

2.2.5 Integrating AoT Sensor Data Connectors with an extensible geospatial science gateway platform (Carol X. Song, Purdue University)

Urban science is taking on many existing and emerging challenges as a result of rapid urbanization and the increasing vulnerability of urban infrastructure and the resident population to extreme weather and environmental changes. While remote sensing, weather, land use, water, etc., data is available on global, regional, and sometimes local scales, urban modeling requires both finer-grained and non-traditional GIS data. Specifically, environmental data on a city-scale with finer temporal and spatial resolutions and near real-time data such as air quality, weather, street level views, noise, movement, etc., will enable researchers to model, analyze and respond to wide-ranging urban challenges. As a geospatial science gateway (MyGeoHub.org), we help researchers utilize and manage the diverse scientific datasets needed in their computational and data analysis tools and collaborate with them to create decision support solutions for data-driven science. In a recent NSF funded data framework project, we are creating an extensible, plug-n-play geospatial data framework to provide seamless connections among platforms, data and tools, making large scientific and social geospatial datasets directly usable in scientific models and applications. A number of our science partner projects are interested in utilizing AoT sensor data to study urban related challenges, including urban flooding, urban design and planning, and risk behaviors during emergency situations, to name a few. Data connectors running on our science gateway software will need to access AoT data through a persistent, reliable and flexible API that can feed custom selections of near real-time sensor data to online computational and analysis tools.

2.2.6 Developing ultra-high resolution air quality forecasting system for urban regions (Rao Kotamarthi, Argonne National Laboratory)

As the amount of data collected increases exponentially, development of suitable numerical modeling tools to assimilate these datasets and developing short-term forecasts of air quality and weather at urban scale is necessary. We propose to build on data assimilation methods developed previously (e.g., Ensemble Adjusted Kalman Filter Methods) in a regional scale framework for WRF-CHEM model to urban scale forecasting. This Bayesian framework for forecasting requires development of prior distributions fields of select trace gases and aerosols using the observational data from AoT observations. We propose to develop forecast models based on AoT data using this framework.

C. Catlett (cec@uchicago.edu)

2.2.7 Evaluating urban representation within weather models for high-impact weather simulations (Dev Niyogi, and Paul Schmid, Purdue University)

Urban areas are becoming highly susceptible to weather extremes. From heat stress to heavy rain-induced flooding, urban areas are becoming the canaries of climate change. Our efforts are currently directed towards developing a gridded representation of urban morphology using different techniques and integrating that information within urban energy balance models. An effort in particular for the World Urban Data Analysis and Portal Tool (WUDAPT) is underway at Purdue which is planning to make the city of Chicago as one of the testbeds for testing the various approaches/algorithms for testing city morphological representation. The key aspect of such representation is not to replicate the google map like imagery rather it is to represent the physiographical characteristics that are required by gridded environmental models that simulate the weather features, energy balance at city block scales. Using two weather models - RAMS and WRF with the WUDAPT input, our goal is to test the performance against AoT data that has been developed and available. The AoT data provides a unique data source for testing such a modeling framework and will be analyzed in details to understand the performance issues and improvements needed in the urban models. We also plan to use vehicular data, and anthropogenic input term for the models and test them using the AoT datasets. These efforts will be likely supported through NSF AGS.

2.2.8 Urban-scale Weather Forecasting over Chicago downtown (Gökhan Sever, Argonne National Laboratory)

Forecasting the weather at urban-scales requires developments of models that are capable of running at a few cubic meters resolution as well as physical representations of the atmosphere that are verified at scales relevant to buildings. Efforts within the scope of the exascale computing project pave the way toward designing and handling large compute and data requirements of such future modeling infrastructure. While these efforts show rapid progression in the recent years, observational data that capture high spatio-temporal pulse of weather with which models are designed upon and verified against are not scaling up with the computational advancements. Collection of temperature, humidity, and pressure measurements by dense network of the Array-of-Things sensors is an essential step for validating model results within Chicago metropolitan area. Synergistic efforts between the computational and observational fronts would make the forecasting of weather a reality at building scales within the next decade.

2.2.9 Developing sound quality metrics for urban areas (Hales Swift, Ralph Muehleisen, Argonne National Laboratory)

Sound quality metrics are mathematical models of how people perceive and experience sound and model and predict such aspects as loudness, sharpness, roughness, tonal character or other specific perceptual attributes of sound. As such, they allow researchers to more accurately account for the impacts of sounds in such socially relevant categories as annoyance and stress. Most sound quality metrics are mathematically fairly straightforward and can be implemented on desktop or laptop grade computing with near-real-time performance. If the AoT GPU-enabled edge processors were employed with an appropriate implementation, real time performance would be easily attainable. Once metrics are calculated they can be transmitted using minimal amounts of data in order to allow characterization of the loudness or other attributes of a space throughout the day. Including loudness metric values would be a benefit to, for example, epidemiological studies on the effects of noise on health since loudness is likely more representative of human impacts of sound magnitude than conventional metrics such as A-weighting.

2.2.10 Measuring the Benefits of Ecosystem Services (Sybil Derrible, University of Illinois-Chicago)

Cities throughout the United States are implementing green infrastructure strategies as a measure to deal with storm water runoff during heavy rains. From simple rain gardens located next to buildings to large low-impact development strategies covering larger areas, green infrastructure is known to generate a slew of benefits in addition to its positive hydrological impacts. For example, air temperature around green infrastructure may decrease in the summer, making the area more comfortable. Some studies also report increase in general well-being thanks to the presence of greenery. Using AoT nodes, we will be able to directly measure which green infrastructure design are able to decrease air temperature. Similarly, we will be able to compare foot traffic (both the number of people and the walking speed) from the cameras in AoT nodes that can be used as a proxy for well-being. These are only two examples, AoT unleashes a wide array of possibilities that have the potential to transform our understanding of ecosystem services.

2.2.11 A Novel Earth Science Curriculum (Max Berkelhammer, University of Illinois-Chicago)

The perception that geoscientists merely study rocks limits recruitment into the geosciences, a field that encompasses a wide array of topics including climate change, air pollution, and sustainable resource management. The discipline increasingly demands skill sets such as computer modeling, big data analysis, and quantitative field and laboratory measurements. Moreover, the field must diversify its cohorts, as geology now is as much about cities and urban runoff as it is about mountains and streams. Through an introductory level class called “Climate, Contamination and Chicago” at UIC, students study the history of Chicago through the lens of a geologist. Students take tours of the city, utilize historical maps, and analyze soil mercury contamination and air pollution data. They are challenged to see their local environment as part of the larger Earth System that encompasses the hydrosphere, biosphere, atmosphere, and geosphere. Through a Course Hero-Woodrow Wilson Fellowship, a novel syllabus is being developed with a focus on the urban atmosphere. Students will utilize new technology for atmospheric measurements, empowering them to address topical issues through fundamental science. AoT data provides an opportunity for analysis of authentic data towards the goal of understanding the factors that influence air quality. Beyond analyzing this data, the students will take part in lab exercises to develop and deploy their own sensors using low cost micro-processors. This helps students develop an understanding of how sensors work so data can be understood as more than numbers coming out of a “black box”. Although it is a Chicago-centric curriculum, the approaches are broadly applicable and the atmospheric science is universal.

2.2.12 Illinois CURES and the Array of Things (Donald J. Wuebbles , UIUC) (team leader for over 20 faculty partners in the University of Illinois system)

The University of Illinois is currently developing the Illinois Center for Urban Resilience and Environmental Sustainability (Illinois CURES). The purpose of the University of Illinois CURES is to harness the unique capabilities at the universities and with its partners (e.g., Argonne) to develop and deploy the research, applied learning and education, with public engagement, to generate the capabilities that cities will need to be more livable, prosperous, resilient, and sustainable. Illinois CURES is aimed at helping cities and surrounding communities across a range of sizes, with an initial focus on cities in the Midwest, but our aim is to be much broader by growing the Center’s reach as opportunities arise to be both national and international. The sustainable cities of the future will utilize a variety of technologies to make them “smarter” through an increasingly effective combination of sensors and digital telecommunication networks. They will make more efficient use of physical infrastructure and natural resources through artificial intelligence and data analytics that support strong and healthy economic, social, cultural development. We expect AoT to be a key component in our working with cities to achieve these goals. In addition, the AoT is vital to establishing an innovative framework to diagnose and tackle complex urban challenges through integrating analysis, measurement, and modeling with novel cyberGIS (geographic information science and systems (GIS) based on advanced cyberinfrastructure) and cyberinfrastructure (CI). The unique combination of these capabilities and related scientific expertise at the University of Illinois and its partners will allow us to realize transformative scientific and engineering understanding for developing innovative solutions to many challenging environmental and urban problems.

2.2.13 Fine spatial and temporal interpolation and prediction of pollutants for health risk monitoring (Jane Lin, University of Illinois-Chicago)

Since 1998, the U.S. Environmental Protection Agency (U.S. EPA) AirNow program has provided hourly air quality measurement data and daily forecasts to the public down to the zip code level. AirNow does an excellent job of connecting the public with air quality monitoring data and forecasts. The data source for AirNow is the ambient air quality monitoring data from the U.S. EPA's air monitoring network, which features thousands of high quality instruments across the country. This valuable data provides citizens with the opportunity to change daily activities to reduce exposure to harmful air pollution. However, most of the monitoring sites are clustered around population centers, leaving many areas far from any monitors. Furthermore, ambient monitors give overall readings of the concentration levels at fixed locations and do not capture the significant spatial variations in concentration levels within a few hundred to a thousand meters from a local source. That is a major drawback in using the stationary ambient monitoring measurements.

To address this issue, we have developed a deep learning framework to predict localized hourly PM2.5 concentration levels with AirNow data coupled with local sensing data and other local specific data including the AoT pollutant data and traffic count data as well as land use and roadway network data. The localized PM2.5 predictions combined with user activity and location information will produce so called personalized air quality index (PAQI) by adjusting the air-wide AQI values by individual activity and microenvironment. In the future, when AoT collects in-door air quality data, it will be valuable to improve the prediction of PAQI as people generally spend 60-80% of their time indoors.

2.2.14 Identifying Vulnerable Communities with respect to Air Quality (Debaleena Chattopadhyay, University of Illinois-Chicago)

Pollution exposure assessment at the population level is an established enterprise for environmental scientists and public health officials—but efforts to help individuals monitor and track their personal pollution exposures have just begun to garner research interest. Self-tracking pollution exposure is challenging for several reasons, including current limitations in sensor size, accuracy, and cost, frequent calibration requirements, and that people's daily activities often interfere with data quality in wearable sensing. We are using AoT urban sensing data to help different vulnerable user groups, such as older adults or developing children, monitor their personal pollution exposure. For instance, recent epidemiological studies have shown a positive association between long-term exposure to PM2.5 and ambient noise and mild-cognitive impairment in older adults, a precursor to dementia.

2.2.15 Computational Oncology and Social Justice (G. Elisabeta Marai, University of Illinois-Chicago)

Together with the Feinberg Foundation, Marai has created a publicly available index of social services in the Englewood and West Englewood neighborhoods of Chicago—among the poorest in the city—creating tools for correlating these services with US census demographics and City of Chicago land occupancy statistics. Together with Dr. Robert Winn, the director of the UIC Cancer Center, Marai is creating software to correlate ward-based cancer statistics in the Chicago area with these social services and US census demographics. AoT's hyper-local measurements of air quality in these neighborhoods will be a valuable source of data towards understanding correlations between cancer occurrence and specific wards. Funding for this work comes from NIH and NSF.

2.2.16 Integrating AoT and CyberGIS for Urban Discovery and Innovation (Shaowen Wang, University of Illinois at Urbana-Champaign)

The CyberGIS Center for Advanced Digital and Spatial Studies was established in 2013 as an interdisciplinary center of the University of Illinois at Urbana-Champaign (UIUC) through a partnership among a number of campus units and the cyberGIS community beyond UIUC. CyberGIS Center supports scientific and user communities by offering advanced cyberGIS expertise and resources. CyberGIS Center houses a set of high-performance computers, big data sources, and related services for performing computation- and data-intensive geospatial analysis and problem solving in various research, education, and outreach contexts. Projects at the CyberGIS Center range across a number of domains such as Earth and environmental sciences, health and sustainability in which major research advances depend on innovative cyberGIS capabilities and multi-scale geospatial problem solving. Data from these projects and agencies such as the U.S. EPA and U.S. Geological Survey are made available to broad research and education activities. CyberGIS Center provides access to advanced cyberinfrastructure environments such as the NSF Extreme Science and Engineering Discovery Environment (XSEDE) and the virtual ROGER supercomputer through software and tools that are developed at the Center. Integrating AoT and cyberGIS capabilities provides unpreceded opportunities for urban discovery and innovation. For example, supported by an NSF DIBBs project (NSF 1443080), researchers from the University of Illinois at Urbana-Champaign and the University of Chicago created UrbanFlow – an interactive online environment to enable integration of diverse spatial data from traditional authoritative (e.g., Census, remote sensing) and modern non-authoritative (e.g. Twitter) data sources for urban discovery and innovation[9]. AoT data is expected to boost the capabilities of UrbanFlow for mapping urban dynamics.

4 Round Table Discussions

4.1 Science: Driving the AoT Instrument Operational Priorities and Capabilities Roadmap

The science breakout included roughly 40 representatives from academia, industry, and government. *The round table was led by AoT Co-PIs Daniel Work (Vanderbilt University; Engineering), Pete Beckman (Northwestern University and Argonne National Laboratory; Computer Science), and Kathleen Cagney (University of Chicago; Social Sciences).* The scientific communities represented included experts both within and at the intersection of the following areas:

- **Environment** including weather, climate science, natural disasters, air quality, environmental science.
- **Built Infrastructure** including urban engineering innovation, urban hydrology and storm water management, energy, and multi-modal transportation and mobility.
- **Social Science and Public Health** including health services research, public policy and practice, and behavioral sciences.
- **Computer Science and Sensor Science** including ultra-lower power sensing and computing, gas and pollutant sensors.

4.1.1 Overarching opportunity of the AoT Instrument

Our group discussed the broad set of opportunities that AoT uniquely enables, ranging from community engagement to scholarly production. We discussed the extent to which long-standing and contemporary questions related to urban context can be addressed, and also the way in which AoT can suggest—and test—new hypotheses. We related this to both the conceptual and statistical models employed, the measures, and either the design for the descriptive analyses or the intervention.

Discussions emphasized the *impact* of findings. How can AoT data be used to inform choices or suggest change in urban space? Examples included individual and community-level health status, air and water quality, mobility, noise and green infrastructure. The group was relatively balanced among scholars, policymakers, and those with links to community organizations; although all emphasized different aspects of AoT and it contributions, general consensus was that AoT can deliver salient information and speak to the goals of each constituency.

4.1.2 New Science Enabled by the AoT Instrument

Many important science questions were raised that have the potential to be informed and answered using AoT as an instrument. For example:

- **Do local climate zones exist?** Urban climatologists can categorize the built and natural characteristics of neighborhoods in cities, e.g., amount of vegetation, hardscapes, etc., which are important to inform urban climate models. AoT offers the possibility to detect and document the existence of local climate zones that behave differently based on these characteristics. Such findings are important to further understand the connections between the urban environment and climate, and indeed to improve the design of urban spaces in terms of energy sustainability.

- **How does infrastructure impact individual and public health?** Understanding the role of the built infrastructure and the environment on individual and on public health more broadly is a promising direction for AoT. An illustrative example was provided noting that in Chicago more than two dozen schools are located within 500 ft of a major road or freeway. Recognizing it is unlikely to move roadways or schools, practical interventions include the development of green buffers to improve local air quality. Scientific questions span the optimal design of such buffers (e.g., vegetation type, density) to improve air quality, and how the improvements in local air quality manifest as improvements to the health of the students in those schools and to the proximal population at large.
- **Social science questions.** A fundamental question in social science research relates to racial/ethnic composition and the role of propinquity. Census data can provide information on racial, ethnic, or economic composition at the tract level but these data cannot provide information on the nature of interaction among these groups. The AoT data provided through image analysis has potential to provide fundamental insights into how often these groups interact in close proximity. Another example relates to emotional well-being and the extent to which it is affected by the presence of and exposure to the natural environment, including both the physical environment and the impact of such factors as noise or air pollutants from traffic or other sources. AoT can provide information on the form and content of green space and the quality of natural surroundings. These data can then be linked to individual-level data on emotional health.
- **Do findings and insights in one locale generalize to others?** As data is collected, models built, and hypotheses tested using the AoT instrument in Chicago, AoT deployments in an increasing number of other cities provide the opportunity for comparative studies between and among cities. The fact that the data can be collected with the same instruments but in different regions opens up the potential to ask questions and validate findings that generalize among cities. Indeed, given the phased deployment of nodes in Chicago, each deployment is an opportunity to set up experiments comparing neighborhoods within Chicago.
- **Science of city health monitoring.** Domain scientists in the domains represented in the breakout have high fidelity models today that can be used for short term forecasts regarding the state of the built/natural/social urban environment. However, such models need to be continuously corrected (e.g., via Bayesian methods) with field data to maintain accurate estimates as time progresses. AoT provides the potential for large data streams to continuously update predictions from the models so they accurately track changes of the urban environment over time.

Figure 14: Science input to AoT roadmap. (Daniel Work, Vanderbilt University).

4.1.3 Needs/Wants/Opportunities

The latter portion of our session was devoted to a wish list of sorts for the kinds of information AoT can provide. Our group returned to a discussion of nested contexts--could we, for instance, combine static sensors (e.g., AoT) with mobile or personal sensors to better understand person-level exposure in a particular environment? While the AoT project has received open offers from various Chicago agencies to mount devices on public transit or city management vehicles, this has been beyond the scope of the current project funding. We viewed such integration of fixed and mobile measurements as useful for both public health and social science. We also considered the environmental attributes of the nodes themselves. How might AoT include not only the location of the node but also other metadata to more fully characterize its environment (that is, near a freeway, with a garbage can nearby), and to what extent can we integrate relevant complementary datasets about the environment in which the node is located? For instance, how might the AoT team most effectively combine data in Plenario (open urban data such as crime, inspections, and infrastructure data) with AoT time series and node metadata?

Throughout the course of our session we also discussed the sensors and how they might be improved or enhanced. For instance, nightly calibration or novel analytics may be employed. On a related note, we also touched on the cost of sensors and how low-cost sensors might be supported or validated to make certain data are of high quality.

We again emphasized that different constituencies may have different needs in terms of science. Policymakers may want to gain insight related to traffic congestion or crowd management, while community organizations may want to understand municipal responsiveness to area-level needs (e.g., filling of potholes). Those in the research community may want to document fundamental descriptive results related to the social and physical environment, or test hypotheses that have some form of causal interpretation (and that can be aided by the longitudinal nature of these data). Finally, we recognized the tremendous potential the AoT project creates to support work *between* the constituencies to leverage scientific findings to implement change to improve the context and performance of the urban environment.

4.2 Data: Providing Diverse Data Services for Science, Policy, Education, and the Public

Roughly 30 participants met to discuss AoT data services, bringing a divers set of perspectives associated with research questions that can be asked of the data. Providing data to these different users, enabling them to readily examine and evaluate AoT data streams of interest to their research, has impact on the nature of the application programming interfaces (APIs) provided as well as on the set of data products that could be provided alongside the ever-expanding corpus of instrument data. *The round table was led by AoT PI Charlie Catlett (University of Chicago and Argonne National Laboratory), AoT Data systems lead Vince Forgione (University of Chicago), AoT software developer Sean Shahkarami (UChicago), and former Chicago Chief Data Officer Tom Schenk (KPMG).*

All AoT data is saved and will be retained and available for future use, but as the volume of this data grows it will be important to enable users to examine the nature of the data in order to determine how or whether it might be integrated into their workflow or analysis. Download services enable users to obtain all data measured to date or subsets such as the most recent days, weeks, or months. Tools are also provided at the download site to enable users to "slice and dice" the data, such as extracting

specific time periods or subsets of measurements and/or nodes. These downloadable data sets are updated daily, supporting longitudinal studies.

In addition to the download services, AoT APIs provide a rich set of query functions optimized for “fresh” data (latency averaging 5 minutes). The APIs are designed to support (a) applications such as for navigating based on urban conditions, (b) monitoring conditions over a particular period of time, such as an ongoing event (e.g., a festival, severe storm, or emergency), and (c) streaming data into science gateways and portals for use by specific communities.

When retrieving data, many users will not want a data set in its entirety but in a downsampled or aggregated format that works for their particular research or information gathering needs and their data storage capabilities. Beyond the tools currently in place for working with the download data, users will in many cases prefer different options for downsampling data, such as median hourly (or daily) values or other sampling functions. For instance, many users will want to extract only one or several measurements from a set of nodes, reducing the volume by at least an order of magnitude. The AoT team noted these as input to further refining the download toolkit.

AoT data is formatted using comma-separated values (CSV), and thus can be readily translated into other formats. The use of an open, simple CSV schema has already enabled the integration of data from multiple Waggle projects, including AoT, environmental deployments of Waggle nodes, an NSF-sponsored project using Waggle nodes to measure pollen in Chattanooga, Tennessee, an NSF-sponsored air quality measurement test with commercial street kiosks in New York City, and sub-$100 “microWaggle” devices measuring moisture and air quality in projects such as Northwestern University’s flood modeling effort in the Chatham neighborhood on Chicago’s South Side. The schema is being considered by many potential partner sensor projects, and the simple, common CSV approach allows such measurements to be seamlessly mixed in with AoT data, where any and all tools and analysis methods can be leveraged.

Some users communities will benefit from translation tools, such as to extract a set of AoT measurements in standard formats such as HDF5 or NETCDF.

In addition to the APIs and download services, the science community will benefit from services that enable them to rapidly assess the key aspects of a data set in a way that communicates the most important statistical properties and data quality in order to determine its suitability for particular science investigations. To this end the AoT team sought feedback from the round table participants regarding the concept of data “thumbnails,” analogous to common search paradigms such as internet search results for videos. How might such “at-a-glance” representations be designed for high-volume data such as created by AoT? That is, 2-dimensional heat maps or other visual techniques that capture some aspect of a data set, such as the areas of Chicago that have experienced PM2.5 measurements above a particular threshold over the course of a recent day, week, or month.

Importantly, the discussions recognized the value of basic download and API services and the importance of keeping all historical data, *emphasizing that any downsampling or summary data products will augment, not replace, the full resolution historical data corpus*.

The remote programmability of AoT nodes was also discussed in context that the sensor sampling methods can be modified or augmented. For instance, the computation within the nodes (“edge” computation) could be used to implement different sampling algorithms. In addition to sound pressure in ten octaves, as it currently provided from microphone readings, a science project may want to compute high/median/low values over arbitrary time periods, or to detect sound “events” during which the sound pressure exceeds a threshold for a minimum period of time. This AoT/Waggle capability

presents a valuable capability in that over time the instrument output can be deepened and extended through scientific iteration and as the community of users expands.

Data formats that may be most applicable are highly dependent upon the research question. Some users may want to extract location-specific snapshots of subsets of measurements over a particular time-period at fixed intervals. For instance, capturing the measurements for a 24-hour period of time at a particular location during which an event of interest occurred such as to understand the environmental or urban activity impacts of a fire or public gathering. On the other hand, a user may desire snapshots of a particular type of data (e.g., air pollutant concentration) at multiple locations to gain high-resolution insights into air quality changes in response to weather and local factors such as factories or highways. Finally, a user may desire to stream data and watch it unfold to observe trends, patterns, and abnormalities such as during a known or scheduled event. All of these examples are possible given the current APIs and download tools, and their diversity underscores the importance of the AoT team's strategy to cultivate partnerships with data providers from various communities, ensuring that the AoT APIs support their import requirements.

As with any instrument or facility, the user community will be diverse in terms of their expertise as well as their data requirements. Many users, particularly those accustomed to "big data," will simply download the data and process it locally. Other users, who may not have the expertise or facilities to work with data larger than a few megabytes, will require downsampled, aggregated data sets as described earlier (or tools to extract such data sets). An important emphasis of the AoT project is also a much broader set of users who are not necessarily expert in "big" data or data science. Consequently, in addition to the partnerships with community-specific data providers, it will be important for AoT to provide some level of user support services, and above all to continue the engagement with a diverse user community as this unique type of instrument evolves.

Finally, scientists will require a mechanism for citing data sets used in publications, such as through digital object identifiers (DOI). The discussion recognized that it will be necessary to determine the optimal fashion for assigning DOIs (for instance, would each month? Year? Day? Receive a unique DOI, or could the results of a filter (e.g., PM2.5 in a particular neighborhood over a specific time period) be given a DOI? Irrespective of the scheme used for assigning DOIs, it is anticipated that as algorithms are refined, such as for translating sensor readings (e.g., voltage output of sensor) into measurements (e.g., parts per billion), data may be re-interpreted and republished. In these cases, both the previous and new data sets will require DOIs along with metadata added to the deprecated data sets such as references to the updated data.

4.3 Deployment: Lessons Learned Deploying Research Cyberinfrastructure in the Public Way

This session covered the policies, agreements, processes, public engagement, and interactions necessary for a successful deployment of research cyberinfrastructure in the public way. Reflecting the academic-city partnership at the core of Array of Things, *the discussion was led by Kate Kusiak Galvin, executive director of the Urban Center for Computation and Data at the University of Chicago, and the current (Danielle DuMerer) and former (Brenna Berman) Chief Information Officers and Commissioners of the Department of Innovation and Technology (DoIT) for the City of Chicago. Group participants also included Von Welch, Chair of the AoT Technical Security and Privacy Group,* and several dozen workshop attendees from England, Taiwan, and other localities interested in deploying their own urban sensing projects.

The AoT project in Chicago is one of the largest deployments of scientific cyberinfrastructure in the public way. Most "smart city" research projects involve proofs-of-concept at limited scale and for a limited time, such as in test corridors or smart districts. Most of these projects are also intended to conduct a particular experiment, by the project team, in contrast to providing flexible services and data to the broader science community. With over 100 nodes installed on public infrastructure (primarily street signal light poles) across the City of Chicago, AoT has established an "instrument" with programmable functions for scientists to use, with accommodation for testing new sensor technologies at scales of dozens to hundreds, and with open and free data services designed for both scientists and the general public. As a persistent cyberinfrastructure and in contrast to a discrete experiment, the AoT project has included a comprehensive public engagement effort aimed at establishing partnerships not only with city officials but with community organizations, schools, and residents. The nature of the AoT instrument, in particular with its computer vision and hearing capabilities, also required a careful and transparent process for developing—and operating within—privacy policies and governance processes. The team found no comparable policies or projects upon which to build these mechanisms, thus the importance of documenting the process and the results for use by other projects.

The discussion highlighted four critical components of the AoT deployment in Chicago: Partnerships, Logistics, Transparency, and Public Engagement. Discussion leaders recounted lessons learned during the Chicago implementation, from its beginnings in late 2011 through initial operation in 2018. Throughout, questions and ensuing discussion focused predominantly on how the AoT model can be replicated for deployments in other cities around the world.

There have been a number of papers written regarding the goals, history, and/or approach that AoT has taken, including a peer-reviewed conference proceedings paper [10] and a profile in Harvard Business Review [11].

4.3.1 Partnerships

Because the City of Chicago manages the public way, it was essential to acquire their support and partnership from the very beginning of the project. The first meetings in 2011 between principal investigator Charlie Catlett and Chicago's then-CIO and DoIT commissioner Brett Goldstein established a strong relationship that has helped the project navigate through complex organizational structures in city government, a federal agency, and a private university. Throughout the subsequent years, the CIO has acted as the primary champion and navigator for AoT team among city ordinances and regulations, allowing the researchers at UChicago and Argonne to deploy at scale. The CIO has also regularly organized discussions with other city departments, for instance bringing together over 40 city leaders from a dozen departments in May 2018 to update them on the project and solicit their input on selecting locations for the next deployment phase in late 2018.

Through ongoing discussions with the CIO and other officials, the AoT team sought and received support from the very top of the Chicago government, including Mayor Rahm Emanuel and former Deputy Mayor Steve Koch. Concurrently, the concept of a city-wide cyberinfrastructure for research and education was included as a specific objective within the Chicago Technology Plan [12] released in 2013. That plan established a broad vision for how the city would use technology and data for inclusion, engagement, innovation, and opportunity. Under Initiative 3, covering “next-generation infrastructure,” the plan proposed that Chicago should be a testbed for the development of new and emerging technology—such as AoT—establishing a top-down mandate for DoIT to facilitate the project.

However, beyond the importance of support from the upper echelons of city government it was also necessary to create buy-in from the departments and city employees that would be instrumental for executing the project. The Chicago Department of Transportation (CDOT) regulates the use of the public way, including the city’s street light poles, and their workers would be the people responsible for installing AoT nodes. DoIT facilitated connections with key personnel at CDOT, initiating frequent meetings that began years before the first node was installed. In early meetings the project team brought prototype units to the electricians’ shop, soliciting their input on electrical safety and mounting systems, effectively making the electricians part of the design and test team. This would prove critical later when installation began, as the electricians understood the goals of the project.

Project researchers regularly met with CDOT at their facilities and invited department representatives to visit the University of Chicago as well as the Waggle laboratory at Argonne National Laboratory. AoT and CDOT representatives regularly discussed how the project could support existing departmental goals, such as traffic management and the Vision Zero [13] program to reduce pedestrian fatalities and injuries. Such discussions clarified the return on investment for CDOT’s assistance with the project.

4.3.2 Logistics

These dialogues with CDOT supervisors, electricians, and other operational personnel also informed the design of the nodes, procedures for installation, node locations, and other key features of the deployment. Because of the early and sustained outreach, CDOT personnel felt ownership for the project, to the point where they are empowered to make on-the-fly decisions, such as moving a node to a different street corner because of an obstructed view, at the time of installation.

Beyond CDOT, DoIT also helped the project interface with other city departments about how they could benefit from AoT data. A working group with representatives from city departments was established to share information on how to use AoT, leading to project proposals and suggested node locations that would help support departmental operations. These conversations also informed how AoT data would be distributed, including the development of APIs for integration into existing city data platforms, such as the public-facing OpenGrid and internal databases.

Processes were also put in place to collect node location proposals from across all AoT stakeholders; not only city departments, but also research groups, communities, and residents. Through community meetings (see below) and the AoT website, we opened channels where anyone could make a suggestion for AoT deployment, including the location of nodes, the type of data they collect, and problems that they could potentially help solve. In the second round of node installations, beginning Fall 2018, approximately 15 percent of the node locations are based on community suggestions via the website or direct interactions.

4.3.3 Transparency

As a technology project placing cyberinfrastructure in the public way, there is critical need for transparency, carefully and clearly protecting privacy, and ensuring system security. Residents’ concerns about public data collection were underscored by early press coverage of the project, at times focusing disproportionately on fears about surveillance and consent around the data that nodes would collect. These responses informed the AoT privacy and governance policies and the legal framework reinforcing the formational commitment to transparency, privacy, and open science.

Beginning in 2015, facilitated by the Indiana University Center for Applied Cybersecurity Research (CACR), researchers and attorneys from the city, UChicago, and Argonne collaborated with external privacy experts on the drafting of these policies. The group faced three main challenges in creating these documents. First, while city technology projects typically focus a specific goal, such as the energy and public safety benefits of a “smart streetlight” installation, AoT is a general-purpose and evolving research platform, with a broad range of applications for science, city operations, and community improvement that more difficult to explain to residents. Second, the policy drafters faced a paucity of existing and/or available policies related to the deployment of cyberinfrastructure in public settings. Consequently, the bulk of the policy had to be written from scratch. Third, unlike many other technology policies, there was no mechanism by which people near AoT nodes could “opt in” to the project and its policies, raising the bar for reassuring residents that their rights are being protected.

The primary mission of the privacy policy was to clearly define what AoT would and would not do, explained at a level that would be accessible to Chicago residents. Because AoT is a development platform technically designed for expandability, the privacy policy needed to establish its current operational limits, the external oversight bodies that would review and approve any future changes, and the mechanisms for informing the public annually about new data requests and new capabilities.

After creating initial drafts of the privacy and governance policies, the AoT team partnered with the *Smart Chicago Collaborative*, a local organization specializing in community education and engagement around technology, to manage a period of public comment on the drafts. This effort included community events held in neighborhoods where nodes were scheduled for installation. At these events, AoT leaders presented the goals of the project, the capabilities of the devices, and the approaches taken regarding security, privacy, and location selection. Residents could read, annotate, and ask questions about the draft policies during the meetings or online. The drafts were also posted to the AoT website where electronic comments and questions were also accepted. Project leaders and the IU CACR engaged expert groups such as the American Civil Liberties Union and the Electronic Frontier Foundation for feedback on the drafts. During a six-month open comment period, the AoT team responded to all comments and questions from these channels publicly via the AoT website, and took the feedback into account as the policies were revised and officially implemented. In addition to the policies, the website includes a report on the public engagement process and a compendium of responses to each and every question or concern that was expressed during (or subsequent to) the comment period. The policies call for such a six-month comment period prior to implementing any normative changes to the privacy policy.

While each municipality has its own unique regulatory and cultural environment for privacy, the policy documents created for AoT in Chicago are made available as source material for similar projects to adapt as appropriate. Additionally, for AoT partner projects (those installing AoT nodes in their cities) the contract language used in the agreement between the University of Chicago and the City of Chicago is made available. Notably, AoT partners deploying nodes in other cities enter a contractual agreement with the University of Chicago that includes the requirement that (a) all data will be open and free, and (b) local privacy policies will be identical to, or stronger than, those used in Chicago.

With privacy predicated on good system security, the AoT project has emphasized security in the design of the system. This has included internal technical reviews by cyber security experts at Argonne National Laboratory as well as a six-month engagement with the NSF Cybersecurity Center of Excellence (https://trustedci.org/aot/) to assess the system security and privacy practices.

4.3.4 Public Engagement

The community engagement meetings planned and held during the privacy policy review period also achieved a broader goal of bringing Chicago residents into the AoT project as active participants. Since the project was envisioned from the beginning as a “community technology,” it was important to inform and solicit feedback from a broad variety of community groups, from civic tech organizations to non-profit advocacy groups to residents with little to no technology experience or science education.

Here the AoT project benefitted from not only the commitment to innovation among city officials and leaders, but a related culture and community associated with the use of data and technology for civic improvement. This culture includes regular meetings, such as the weekly *Chi Hack Night*, established in 2013 to bring people together to discuss and practice the use of data and technology to develop understanding and to explore solutions to urban issues. Concurrently, groups such as *ConnectChicago* are dedicated to improving digital literacy in communities with low access to technology. Our partners in community outreach, the Smart Chicago Collaborative (now part of CityTech), helped to both publicize and conduct our community outreach meetings, based on civic user testing principles developed on other projects with city and industry partners.

Even in our earliest meetings, conversations around privacy and data security quickly evolved into residents proposing ideas for the types of problems and data that the AoT project could address. As AoT data becomes increasingly available, future community meetings will also focus on empowering residents to make use of that data, making sure that it is a useful tool for communities and advocacy groups, not just researchers and government agencies.

Public engagement has also taken place through media channels. Since its early stages, AoT has captured the interest of reporters from around the world, and despite some early misrepresentation and negative framing, the vast majority of news coverage has been positive. By being responsive to media requests for interviews and publishing in-depth information about the project on the AoT website, coverage has been more accurate—and more positive—improving public perception and engagement. Indeed, many public requests for node locations cite a news story as their introduction to AoT and its research goals.

4.3.5 Conclusion

As other cities deploy AoT instruments or similar cyberinfrastructure, the AoT team will continue to capture lessons learned in order to deepen the guidance on the important non-technical steps for deployment. Each city’s culture and relationship with technology will be unique, thus the partnerships and processes created in Chicago may not be precisely replicable. But to date it has been evident that the general principles of academic-city partnerships, transparency, and public engagement are essential to the success of any project deploying cyberinfrastructure in the public way.

5 Tutorials / Mini-Workshops

Three concurrent sessions provided both tutorial content and feedback from users. The first (Analyzing AoT Data) featured tools and mechanisms from several AoT science partners. The second (Edge Computing) provided detail (and sought feedback) regarding the machine learning capabilities available for users on the nodes. The third (“School of Things”) covered a curriculum developed with funding from the Motorola Solutions Foundation, summarizing the materials that have to date trained over 400 high-school students in Chicago.

C. Catlett (cec@uchicago.edu)

5.1 Analyzing AoT data (hands-on, tools and frameworks for using AoT data)

In this workshop, participants were exposed to examples of AoT analytics tools for a variety of environments. Users learned to access, wrangle, and analyze AoT sensor data using R and R Markdown as well as Python and Jupyter notebooks. Materials were provided to enable participants to set up their laptops, equipping them to analyze data and modify the materials for their own uses. *An introductory set of example GIS analysis tools was presented by Anand Padmanabhan from the CyberGIS Center for Advanced Digital and Spatial Studies, University of Illinois at Urbana-Champaign.*

Spatial Analytics with R and R Markdown: Tools covered included those demonstrating how to access and clean AoT data, being sure to enable its spatial features. Additional tools were covered such as those used to plot the AoT sensor data in an interactive map, to inspect and explore specific sensor stream such as temperature values, and to generate an interpolated surface of temperature for the City of Chicago. After these guided workbook sessions, participants learned about future applications of AoT data and how these tools can be used for other exploration such as to measure and predict air quality. *The R and R Markdown tutorials, taught by Marynia Kolak and developed by the UChicago Center for Spatial Data Science, are online: https://geodacenter.github.io/aot-workshop/.*

Python Jupyter Notebooks: Participants were introduced to Jupyter Notebooks (www.jupyter.org) and the use of python for the exploration of AoT datasets. Participants were first introduced to the Jupyter Notebooks, how they could be used to develop code, document process and be shared with collaborators. Using python and pandas, participants were then shown how to locate data and load the data into a pandas dataframe for manipulation. Participants initial activities were to load the nodes.csv dataset, where they learned how to view the data, plot a histogram of when AoT nodes began to report data, plot the cumulative number of nodes deployed in Chicago. The participants then learned how to use the latitude and longitude data included in the ndoes.csv file to produce a map of Chicago with AoT locations plotted. Finally, the participants were able to take information on nodes that were not producing data anymore to create a map of active and inactive nodes in Chicago. Along the way, via mapping the participants were able to find a node outside the city that was part of the dataset.

Participants were then shown how to load sensor.csv and browse that data before moving the large data.csv dataset. Given load times and sheer size of the data.csv a quick demonstration of how to filter out smaller datasets was done, producing a single node dataset that spanned the entire reporting period of the given node and a second dataset that contained a week's worth of data for all Chicago AoT nodes. Based on these two datasets participants were taught how to convert data components to plottable values, convert date and time information from UTC to the central time zone and make it the index for smaller datasets. Participants were then able to plot relative humidity for a single node, learned how to dive into the details, adjusting scale as needed. Participants plotted changes in light intensity captured by the node. As a final exercise the participants plotted the temperature from all nodes to see how close they were to each other.

The overall goal was to provide some simple and clear examples how individuals interested in working with AoT data could get started and provide a starting point for their own explorations. *The Python Jupyter Notebooks tutorial was developed by the Data, Devices, and Interaction Laboratory (ddiLab), directed by AoT Co-PI Michael Papka at Northern Illinois University.* All materials are available at the Community Sensing GitHub Repository (https://github.com/ddiLab/CommunitySensing).

5.2 Using AoT's Waggle Platform for Edge Computing and Machine Learning

Roughly 25 participants learned the basics of machine learning approaches to computer vision and the software environment supported by the Waggle platform. Much of the mini-workshop involved interactions with users and prospective users of the computer vision capabilities, with the Waggle team gathering feedback on the types of "observations" that are of interest for various disciplines. Examples of observations that are planned for late 2018 in AoT include flood / standing water detection, regular count of pedestrians and vehicles (buses, trucks, cars, bicycles), as well as safety-related measurements such as detecting the use of helmets for cyclists. An additional example from an Argonne research project with the U.S. Air Force was also included, demonstrating the use of the sky-facing cameras in AoT nodes to detect drones (noting that the movements of planes, drones, and birds are key to identification rather than the particular shapes). The workshop discussed issues and plans regarding the computational and memory capacity of the current nodes, targeting more resource-intensive observations such as detecting pedestrians crossing against a light (a measure of safety) or vehicles rapidly decelerating (e.g., a "near miss.").

The tutorial also covered the audio and Images pipeline used for gathering sample images, and an overview of how these can be used for algorithm training. This included reviewing example images and discussing challenges such as annotation (given that the AoT privacy policy requires signed data use agreements for access to these training images, thus they cannot, within the present policy, be placed in crowdsourced annotation services. For support of audio analysis, audio samples were reviewed (sampling rate, duration, microphone frequency response and sampling limits, etc.).

Finally, participants engaged in a discussion regarding upgrades planned for AoT nodes in terms of computational support. This included a comprehensive survey of emerging edge (machine learning, distributed learning, AI) hardware, with a presentation including updates from HotChips 2018.

This tutorial was taught by AoT Co-PI Pete Beckman (Northwestern University and Argonne National Laboratory), Nicola Ferrier (University of Chicago and Argonne National Laboratory), and Rajesh Sankaran (Argonne National Laboratory)

5.3 "School of Things" ("SofT") High School and Middle School Curriculum

Figure 15: Lane Tech High School students preparing sensors for installation at Wrigley Field. Photo: Chicago Tribune.

A dozen participants attended the education mini-tutorial, which discussed a curriculum developed by AoT partners with funding from the Motorola Solutions Foundation. The origin of the curriculum was largely from public engagement meetings held by the AoT team early in the project, which generated considerable interest from high school students as well as industry partners, and from parents who urged the AoT team to engage youth. In part motivated by a new Chicago Public Schools computer science requirement for all high school students starting with the class of 2020, UChicago formed an education team with AoT partners from the School of the Art Institute of Chicago (SAIC) and computer science teachers from Chicago's Lane

Technical High School to explore a hands-on computer science curriculum. The resulting “Lane of Things” (LofT) program has trained over 400 students in 2016, 2017, and 2018.

Funded by the Motorola Solutions Foundation and in-kind contributions from component providers, LofT uses low-cost “Internet of Things” technologies to teach students to consider a science question, develop an approach to measuring relevant data, design and assemble a wireless sensor device, deploy the device for measurements, and analyze the data in the context of their science question.

During this tutorial, participants were given a summary of the 5-day workshop the SofT team developed and taught to Chicago high-school and middle-school teachers in July 2018 as part of an expansion of the curriculum to twelve partner schools. Below is a summary of the content.

After the first delivery of the eight-week curriculum to 150 students in 2016 the team reviewed and updated the program for 2017. The 2017 curriculum taught 155 students and 145 are enrolled in the 2018 course that began in early February. In three years, several core approaches have been refined and improved. Students work in three-person teams to develop their own research questions and sensor node designs, choosing everything from the material used for the enclosure to the types of sensors used. These teams install their sensor nodes around the school and use online tools to analyze the collected data, prompting new ideas about using technology to improve their surroundings. All of the projects are then presented at a final event at the end of the curriculum, where teams present poster sessions and demonstrations for scientists and policy makers from throughout the Chicago area, including AoT leaders and partners.

Two improvements were made in 2017. First was the inclusion of an introduction to concepts in urban surveying, device design, and sensing, stepping the students through an introductory mini-project where they learned about the fundamentals of sensor device making with low cost and open source hardware and software systems. This proved to be an important improvement over the 2016 curriculum in that it familiarized the students with the diverse components and concepts covered in this curriculum, including a better introduction to data analysis and visualization. Second was a new requirement that the teams document their projects online—including the “story” of the project’s goals and how things went, all parts and materials, code, wiring diagrams, and instructions for any components made with tools in the makerspace (e.g., 3D print files).

In the first two years, students were encouraged to consider experiments and topics related to their own science interests. In 2018 the student teams were instructed to focus rather on questions and corresponding measurement projects associated with the interests of an organization (that is, requiring that students understand and translate the interests of others). Through a partnership established by the AoT team with the Chicago Cubs baseball organization, students were tasked with developing experiments aimed at questions related to the organization’s venue: Wrigley Field. Student teams were given the choice of developing projects aimed at understanding either (a) venue comfort, (b) venue noise, or (c) rapid customer feedback. To this end, students developed either a weather sensor, sound sensor, or “sentiment” sensor (a unit with buttons to select to answer a Yes/No question as fans exit or enter the venue. Students were able to install their projects in and around Wrigley Field, giving them experience in translating policy questions into science questions.

The LofT team formalized the curriculum for replication, renaming it “School of Things,” (SofT). Each stage of the curriculum has been broken into modules, increasing the flexibility that participating teachers have to pick and choose the topics that best fit their classrooms. These modules are interlaced with the foundational goal of teaching students to explore relevant metrics and how their sensing projects might provide related insights, targeting those areas where sensor systems are well-suited, such as air quality, greenhouse gas reduction, or flood mitigation. The modules include:

1. Overview of AoT, introducing the larger picture of the opportunities with technology and data
2. Microcontroller basics
3. Wiring simple sensors; I/O
4. Potential sensor options, code directory (highlighting a number of sensors), and relationships between problems and sensor solutions
5. Enclosures and prototyping a sensor box
6. Outfitting enclosure, making it a working prototype, and deployment
7. Data analytics and visualization to understand environmental effects and policy

In 2018-2019 the multi-school program will focus students on projects to be deployed in the Chicago Museum of Science and Industry. The plan under development is to enable students to design wireless sensing devices to place throughout the museum to serve both as experiments for the students and as a highlighted exhibit for museum visitors.

This tutorial was taught by AoT Senior Personnel Douglas Pancoast from the School of the Art Institute of Chicago and Kate Kusiak Galvin, Executive Director, Urban Center for Computation and Data at the University of Chicago.

6 Workshop Agenda

Array of Things User Workshop
29-30 August 2018
Venue: Argonne National Laboratory

——- 29-August ——-

0900-0915 Welcome and Workshop Goals (C. Catlett, UChicago/ANL)

0915-1045 State of the Array of Things: Instrument and Platform
Deployment, Example Use Cases, Status (Charlie Catlett, UChicago/ANL)
Waggle Platform Architecture and Status (Pete Beckman, Northwestern/ANL)

1100-1200 State of the Array of Things: Science and Data
Preliminary Evaluation of Air Quality Sensors (Mark Potosnak, DePaul)
Current and Planned Data Pipeline, Format, and APIs (Vince Forgione, UChicago)

1200-1300 Lunch Breakouts for In-Depth Status and Futures Discussions
Waggle Platform and Edge Computing (Pete Beckman, Northwestern/ANL; Raj Sankaran, ANL; Nicola Ferrier, UChicago/ANL)
Air Quality Evaluation Strategies (M. Potosnak, DePaul; R. Kotamarthi, ANL)

1300-1415 Application Lightening Talks
(8-10-minute talks on use, or planned use, of AoT data and edge computing)

- *Urban Microscale Modeling and Developing Heat Mitigation Strategies (Ashish Sharma, University of Notre Dame)*
- *Understanding Environmental Effects on Behavior (Marc Berman, UChicago)*
- *Transportation and Environmental Research (Nicola Ferrier, UChicago/ANL)*
- *Neighborhood-Resolution Environmental Data for Aging Studies (Kate Cagney, UChicago)*
- *AoT and Waggle to Support Atmospheric Sciences (Cristina Negri, UChicago/ANL)*
- *Using AoT for Transportation Research (Dan Work, Vanderbilt University)*

- *Integrating Environmental and Socioeconomic Data with Electronic Medical Records and Molecular Phenotypes (David Liebovitz, UChicago)*
- *Partnership for Healthy Cities (Raed Mansour, Chicago Department of Public Health and Marynia Kolak, UChicago)*
- *Distributed High-Frequency Urban Sensing for Flood Prediction and Impact Assessment (Aaron Packman, Northwestern University)*
- *microWaggle for Hydrology (Vivien Rivera, Northwestern University)*

1545-1700 Toward a User-Driven Instrument: Round Table Strategy Breakouts

Science: Instrument Science and Capabilities Roadmap
(Dan Work, Vanderbilt; Pete Beckman; Northwestern/ANL; Kathleen Cagney, UChicago)

Data Services Outputs
(Charlie Catlett, UChicago/ANL; Tom Schenk Jr., KPMG (formerly City of Chicago); Vince Forgione, UChicago, Sean Shahkarami, UChicago)

Deploying Cyberinfrastructure in the Public Way
(Kate Kusiak Galvin, UChicago; Danielle DuMerer, City of Chicago; Brenna Berman, CityTech (formerly City of Chicago), Von Welch, Indiana University)

—- Evening reception - Argonne Guest House Patio —-

——- 30-August ——-

0900-0945 Breakout Reports and Discussion
1000-1230 Mini-Workshops (Session 1)
1230-1330 Lunch (workshops continue)
1330 Adjourn

Mini-Workshops

A: Analyzing AoT data (data-tools, hands on with spatial analytics, Jupyter notebooks)
Anand Padmanabhan and Haozhi Pan, UIUC Cyberinfrastructure and Geospatial Information Laboratory
Marynia Kolak, UChicago Center for Spatial Data Science
Michael Papka, ANL/NIU

B: AoT Edge Computing and Machine Learning
Pete Beckman, Northwestern/NIU
Nicola Ferrier, UChicago/ANL
Rajesh Sankaran, ANL

C: “School of Things” (“SofT”) High School and Middle School Curriculum
Douglas Pancoast, School of the Art Institute of Chicago
Kate Kusiak Galvin, UChicago

Figure 16: AoT User Workshop Group Photo. Source; Argonne National Laboratory.

7 Participants

Over 100 participants came from across the United States as well as from Canada, Taiwan, and the United Kingdom. While many participants could be classified in multiple roles, a rough approximation is as follows:

60% Scientists and students using or intending to use the instrument, including both data users and scientists interested in programming the instrument (e.g., computer scientists with interest in computer vision, resilient systems, networking, etc.).
15% AoT design/technology team and partners.
15% Research institutions involved in deploying or planning to deploy AoT instruments in their cities.
10% City of Chicago deployment partners.
10% Other Chicago stakeholders (community groups, government, etc.).

(NSF MRI grant Co-Principal Investigators are shown in **bold text;** speakers are shown in *italics)*

Last Name	First Name	Institution
Baddeley	Michael Graham	University of Bristol
Barbour	William	Vanderbilt University
Bartley	Dennis	Argonne National Laboratory
Beckman	**Peter**	Argonne National Laboratory / Northwestern University
Berman	*Marc*	University of Chicago
Berman	*Brenna*	CityTech (UILabs)
Bettencourt	Luis	University of Chicago
Brown	Demetrius	Americorp Vista-GECDC
Brown	Maxine	University of Illinois-Chicago
Cagney	**Kate**	University of Chicago
Carhart	David	Astrotech
Catlett	**Charlie**	Argonne National Laboratory / University of Chicago
Chattopadhyay	Debaleena	University of Illinois-Chicago
Chen	Yuxin	University of Chicago
Choe	Kyoung	University of Chicago

Cohen	Ellen	University of Chicago
Craig	Brian	Argonne National Laboratory
Crowley	Ada	University of Chicago
Cruz	Isabel	UIC
Derrible	Sybil	University of Illinois-Chicago
DuMerer	*Danielle*	City of Chicago
Dunn	Jennifer	Northwestern University
Emmanuel	Abraham	City of Chicago (Department of Transportation)
Etlinger	David	Arup
Ferrier	*Nicola*	Argonne National Laboratory
Forgione	*Vince*	University of Chicago
Fuoco	Marta	U.S. Environmental Protection Agency
Garcia	Cynthia	University of Illinois-Chicago
Garg	Nikhil	CSIRO (Australia)
Gloudemans	Derek	Vanderbilt University
Gunner	Sam	University of Bristol
Hall	Carrie	Illinois Institute of Technology
Hernandez Gonzalez	Liliana	Northwestern University
Hurdelbrink	Douglas	City of Chicago
Jacob	Robert	Argonne National Laboratory
Jacobson	Peter	Panasonic
Jain	Raj	Argonne National Laboratory
Jiang	Junchen	University of Chicago
Kaberon	Dan	Dell (retired)
Keahey	Kate	Argonne National Laboratory
Kolak	*Marynia*	University of Chicago
Kotamarthi	Rao	Argonne National Laboratory
Kusiak Galvin	*Kate*	University of Chicago
Levy	Jonathan	City of Chicago
Lewis	Kimberly	University of Chicago
Leynes	Gene Walker	City of Chicago
Liang	Haorui	The State University of Rutgers
Liao	I-En	National Chung Hsing University (Taiwan)
Liebovitz	*David*	University of Chicago
Lin	Fang-Pang	NARLabs (Taiwan)
Lucius	Nicholas	City of Chicago (Dept of Innovation and Technology)
Mansour	*Raed*	City of Chicago (Dept of Public Health)
Martin	Tami	Argonne National Laboratory
Michaud	Daniel	ARUP
Michelson	Dave	University of British Columbia
Mitchum	Rob	University of Chicago
Mudd	Susan	Environmental Law & Policy Center
Muehleissen	Ralph	Argonne National Laboratory
Mydlarz	Charlie	New York University
Negri	*Cristina*	Argonne National Laboratory
Niyogi	Dev	Purdue University
Oikononou	George	University of Bristol
Ollerer	Jo Ann	City of Chicago
Olopade	Sola	University of Chicago
Packman	*Aaron*	Northwestern University
Padmanabhan	*Anand*	University of Illinois-Urbana-Champaign
Pan	*Haozhi*	University of Illinois-Urbana-Champaign
Pancoast	*Douglas*	School of the Art Institute of Chicago
Papka	**Michae**l	Argonne National Laboratory / Northern Illinois University

Parham	Jamila	City of Chicago
Peaslee	David	SPEC Sensors
Peterson	Norm	Argonne National Laboratory
Philpott	Steven	XtraMedium
Potosnak	*Mark*	DePaul University
Rivera	*Vivien*	Northwestern University
Sakhnini	Nina	University of Illinois-Chicago
Samal	Chimaya	Vanderbilt University
Sampat	Sanil	University of Chicago
Sankaran	*Rajesh*	Argonne National Laboratory
Scharg	Ari	Edelson
Schenk	*Tom*	KPMG
Schmid	Paul	Purdue University
Sever	Gokhan	Argonne National Laboratory
Shahkarami	*Sean*	University of Chicago
Sharma	*Ashish*	University of Notre Dame
Silva	Mariana	University of Notre Dame
Small	Jason	City of Chicago
Song	Xiaohui Carol	Purdue University
Spenko	Matthew	Illinois Institute of Technology
Stetter	Ed	SPEC Sensors
Sullivan	Ryan	Argonne National Laboratory
Tolva	John	CityFi
Tryfonas	Theo	University of Bristol
Weigert	Karen	The Chicago Council on Global Affairs
Welch	*Von*	Indiana University
Work	**Daniel**	Vanderbilt University
Wuebbles	Don	University of Illinois-Urbana-Champaign
Yan	Eugene	Argonne National Laboratory
Yang	Yongchao	Argonne National Laboratory
Yu	Ja Eun	University of Illinois-Chicago
Zachos	Lee	Argonne National Laboratory
Zheng	Heather	University of Chicago

8 Background Reading about the Array of Things

8.1 Links

Array of Things Project Overview, Objectives, and Lessons Learned (2017)

Waggle Platform (2016)

AoT Project Website

High School Curriculum Program (2017)

AoT Data and Tools

The **R and R Markdown tutorials**, taught by Marynia Kolak and developed by the University of Chicago Center for Spatial Data Science, are online:

https://geodacenter.github.io/aot-workshop/.

The **Python Jupyter Notebooks** tutorial, taught by AoT Co-PI Michael Papka (NIU and Argonne) and developed by the Northern Illinois University Data, Devices, and Interaction Laboratory (ddiLab), are available at the Community Sensing GitHub Repository:

https://github.com/ddiLab/CommunitySensing

8.2 Additional References and Reading

[1] C. E. Catlett, P. H. Beckman, K. A. Cagney, D. B. Work, and M. Papka, "MRI: Development of an Urban-Scale Instrument for Interdisciplinary Research (NSF 1532133)," ed. Chicago, Illinois: National Science Foundation, 2015.

[2] Multiple. (2018). *Urban Center for Computation and Data — Past Events*. Available: http://www.urbanccd.org/past-events

[3] P. Beckman, R. Sankaran, C. E. Catlett, N. Ferrier, R. Jacob, and M. Papka, "Waggle: An open sensor platform for edge computing," in *SENSORS, 2016 IEEE*, 2016, pp. 1-3: IEEE.

[4] C. E. Catlett, P. Beckman, and J. Tolva, "AoT Kickoff Partner Workshop," in "UrbanCCD Workshops," University of Chicago, Chicago, Illinois2015, Available: http://www.urbanccd.org/s/Partner-Kick-Off-Report.docx.

[5] C. E. Catlett *et al.*, "Plenario: An Open Data Discovery and Exploration Platform for Urban Science," *IEEE Data Eng. Bull.,* vol. 37, no. 4, pp. 27-42, 2014.

[6] C. E. Catlett, "EAGER: Prototyping an Urban Data Cyberinfrastructure for Computational Social Sciences (NSF-1348865) ", ed. Chicago, IL: National Science Foundation, 2013.

[7] C. E. Catlett and V. Forgione. (2018). *Plenario API to AoT Data*. Available: https://plenario.docs.apiary.io/#

[8] M. L. B. Habibul Ahsan. (2016, 5/15/18). *ChicagO Multiethnic Prevention And Surveillance Study (COMPASS)* Available: http://compass.uchicago.edu/page/about-study

[9] A. Soliman, K. Soltani, J. Yin, A. Padmanabhan, and S. Wang, "Social sensing of urban land use based on analysis of Twitter users' mobility patterns," *PloS one,* vol. 12, no. 7, p. e0181657, 2017.

[10] C. E. Catlett, P. H. Beckman, R. Sankaran, and K. K. Galvin, "Array of things: a scientific research instrument in the public way: platform design and early lessons learned," in *Proceedings of the 2nd International Workshop on Science of Smart City Operations and Platforms Engineering*, 2017, pp. 26-33: ACM.

[11] R. Lal and S. Johnson, "Chicago and the Array of Things: A Fitness Tracker for the City," *Harvard Business Review,* Case Study March 2017 2017.

[12] C. o. Chicago, "Chicago Tech Plan," City of Chicago, Chicago2013, Available: https://techplan.cityofchicago.org.

[13] C. o. Chicago. (2016). *Vision Zero Chicago*. Available: http://visionzerochicago.org

C. Catlett (cec@uchicago.edu)