Collective Sensor Networks: Structures, Outcomes, Promises and Shortcomings

Martin Dittus, UCL Centre for Advanced Spatial Analysis

29th March 2012

A growing number of Internet communities are experimenting with new collaborative sensing practices, where distributed groups take environmental measurements with electronic sensors and publish them online in central repositories. This may include sensor readings of temperature, humidity, home energy usage; but also of radiation levels, indoor and outdoor air quality levels, and many other kinds of sensor data. Kera and Graham have given such collaborative efforts the name *"Collective Sensor Networks*" (CSN.) [1]

The Weather Underground [2] may be one of the most successful projects so far, with a clear focus and supported by a strong founding team of experts. It sells weather station sensors to a global community, aggregates their data and uses it to provide weather maps and a weather forecast. OpenEnergyMonitor [3] is building open source energy monitoring hardware and software, and many of its participants publish their personal measurements online. And a small group of experts and enthusiasts is currently designing an "AirQualityEgg", an easy-to-use consumer device that will take air quality measurements and publish them online in a central repository. [4] There are countless more examples, and it has become hard to keep track of all new initiatives.

Collective sensor networks are appealing in a number of ways. Compared to more centralised systems they may lead to a distribution of infrastructure cost, increase the number of observers, improve infrastructure redundancy, and more; but they can also offer new perspectives. In December 2011 the Australian technologist Andrew Fisher described one desired outcome of CSN activity by introducing the term "*Sensor Commons*":

> For me the Sensor Commons is a future state whereby we have data available to us, in real time, from a multitude of sensors that are relatively similar in design and method of data acquisition and that data is freely available whether as a data set or by API to use in whatever fashion they like. [5]

While at this early stage it is hard to separate mere short-term experiments from practices that will demonstrate long-term merit, it is possible to highlight a selection of current sensing practices, and attempt to provide some early insights. What are the motivations to contribute to such systems? How are such communities organised, what level of explicit coordination is necessary? And what are the current experiences and insights that will shape future practice?

In this essay I will first review conceptual frameworks that describe such systems, then introduce exemplary radiation sensing projects that emerged after the Fukushima Daiichi nuclear catastrophe, and provide an initial evaluation of their efforts based purely on primary and secondary sources of documentation, with a focus on project structures and outcomes.

Conceptual Frameworks

Before some exemplary CNS projects are evaluated below the following sections will first introduce a description of CSN systems, some theories of understanding CSN activity, and finally methods of evaluating CSN structures and outcomes.

The System: Building Blocks for CSNs

In order to be able to describe and compare CSN projects it is necessary to understand their constituent elements. Most importantly each of these projects needs a definable scope: the notion of an *environment or resource* that is to be monitored, together with a statement of the *goals and motivations*. This may include a particular geographic scope; an example is the measurement of air quality in a particular area of a city.

Another aspect is the project *infrastructure*: its sensors, data stores, tools for analysis and visualisation, and presentation platforms. This entails questions of system cost, and relates to the project's level of focus (from single-purpose infrastructure to usage-agnostic platforms.)

Along with this there needs to be a description of its *participants*: the project contributors, but also its audience; and an understanding of the various communities that are formed out of any combination of these agents.

Furthermore such a project will entail a number of *practices*: the ways in which data is collected, aggregated, analysed, presented and consumed. Any of these can be a communal or solitary activity. This also entails an understanding of the *social dynamics* of the project: who provides the incentive to contribute? Is the project driven by a central (commercial, scientific, or volunteer-driven) body, or by a heterogenous community of self-interested agents? Do the motivations to participate align with the aims of rigorous data collection? Who owns (plans, funds, builds and maintains) the sensing network?

In 2008, Vinyals etc al introduced a "taxonomy for sensor networks" [6] that describes four feature groups: *sensors*, *networks*, *environments* and *goals*; since they focus on centrally designed censor networks this model does not incorporate participating agents and their social dynamics, and any practices necessary to sustain a project's activities are only mentioned implicitly. In the same year, Gouveia et al introduced a framework for Environmental Collaborative Monitoring Networks (ECMN) [7] with three core building blocks: *motivated citizens*, *sensing devices*, and *back-end information infrastructure*. While they do consider the motivation of participants they only implicitly describe some of the social dynamics, and the framework does not attempt to classify the monitored environment or the types of project aims.

Theories of Understanding CSN Activity

There is little existing literature specifically relating to CSN practices, but there is a wide range of literature and conceptual models that address questions of collaborative data gathering and analysis practices, information access, and modes of participation; particularly in the field of Geographic Information Science (GIS.) The terminology for such systems is equally rich: Volunteered Geographic Information (VGI), [8] Participative GIS, [9] GIS/2, [10] User-generated Content, [11] and more. Within limitations a lot of these can be applied to CSN activities. Existing research on Citizen Science addresses questions of scientific rigour.[12]

Motivations to participate in unpaid, volunteer-driven cartographic efforts are fairly well studied in the context of VGI. Goodchild suggests self-promotion and the desire to share insights with others as motivations to contribute to OpenStreetMap efforts. [8] Välimäki suggests a desire to improve existing mapping data, particularly that of the own geographic environment; the opportunity to participate in a unique and novel activity; the enjoyment of the activity itself; and other factors. [13] Motivations to contribute to CSN projects are likely to a mixture of the above, along with a desire to fill collective information needs that are not yet sufficiently addressed.

In understanding *organisational structures and social dynamics* of CSN projects one can draw from a wealth of existing research in the social sciences, including Mintzberg's models of organisational structure, [14] Small Group Research, [15] Collective Action as introduced by Olson, [16] and collective action in the context of common-pool resources as studied by Ostrom. [17] These works investigate decision-making structures, levels of cooperation, abilities to draw boundaries around groups of contributors, abilities to execute graduated sanctions for the violation of community rules, and many other related matters.

When discussing the viability of CSN project structures Fisher described five fundamental properties that he deemed necessary: the gaining of trust of its participants and audience; the ability to be dispersible, to widely spread sensor installations and participation; the state of being highly visible, both in activity and in outcomes; openness in choice and development of platforms (he mentions open source licenses for software and hardware); and the ability to be upgradeable, to make use of new technical developments as a project matures. [5]

Methods of Evaluating CSN Structures and Outcomes

There are a number CSN project outcomes that can be quantified, including the volume of published data, the period over which sensing activity could be sustained, and the geographic areas covered by this sensing activity; this can be related to the initial project aims as a means of *evaluating project success*.

Comparative studies can help evaluate data quality, provided high-quality reference data is available. In its absence it is possible to simply evaluate internal consistency by e.g. determining the statistical variance of contributions within a spatial region. It is also worth evaluating whether data quality is improved by drawing from a larger pool of contributors, as Haklay et al have demonstrated for OpenStreetMap data. [18]

Finally it is necessary to evaluate operational aspects of a project, e.g. the degree to which collaboration succeeds, and the degree to which the infrastructure used can accommodate project needs. This can be done in *case studies* based on interviews, focus groups, by reviewing documentation, and observing project activity in other ways.

Exemplary CSN Projects

On 11 March 2011 the nuclear reactor facility Fukushima Daiichi experienced a series of failures, which resulted in the release of large volumes of radioactive material. Initially Japanese officials appeared to underestimate the scope of the failures, and the Tokyo Electric Power Company (TEPCO) failed to provide frequent status updates to the public. At the end of March Greenpeace published ad hoc radiation sensor readings of the area, along with some criticism of the government's evacuation plans, which were declared too limited in scope. [19]

There was growing public distrust in official reports, and a strong public desire to establish clarity. Shortly after the event some individuals started aggregating public information resources, but also taking their own sensor measurements and publishing them online. Out of this emerged a number of volunteer groups that attempted to coordinate such efforts.

Pachube Community

The London-based Internet company Pachube offers an online platform for the publication of real-time sensor data, along with and API and data visualisation and exploration tools, and is host to a growing community of sensor data enthusiasts. In the wake of the Fukushima disaster this community started coordinating, creating public awareness, and educating others on how to set up their own radiation sensors. [20][21][22] A week after the nuclear accident Pachube decided to offer free accounts to all users publishing radiation sensor feeds to the public. [23] In June 2011 Pachube announced that they now hosted 2,500 real-time radiation sensor feeds set up by its community. [24] While the initial focus was on publishing raw sensor data, later individuals started producing maps, data visualisations, and other applications. [25][26]

The use of Pachube as central aggregator offered a number of benefits. The cost of the sensing activity was spread across a community of participants. Their feeds were updated in real-time, and available to the public. Data could come from a number of sources, including data published on other websites; it was possible to search and explore their repository of feeds, and data could be exported in a consistent format.

However the heterogeneous selection of hardware and the inconsistency in sensor setups and calibration practices prompted discussions about the quality of the data and the role of scientific rigour and consistency. In addressing this, Pachube employee Ed Borden raised the question: "Is it more useful to know if the value at a particular location is exactly '.075 microsieverts' or if it has been steadily rising over the past 3 days?" [24] Additionally much of the activity was not sustained over long periods. As of January 2012, only around 200 radiation sensor feeds remain active. [27]

Marian Steinbach

The user experience expert Marian Steinbach was among the first individuals to aggregate public radiation sensor information from existing sources. While the official data sets he found had been collected by experts and were regarded to be of a high quality, historic data was only provided in the form of graphs, and it was hard to extract raw sample data. [28]

Initially Steinbach manually captured this data in a public Google Spreadsheet document; later this process was automated. He also invited others to contribute, and allowed the submission of manual measurements. Some of this data was then published on Pachube. From March to November 2011 seven million records were captured, and remain downloadable in a CSV format. Various applications and dashboards were built by Steinbach and others to present the data, [29] and the data was itself aggregated by participants of other collaborative projects. [30]

While this makes for an encouraging example of an individual being able to improve public access of environmental sensor data with limited means, and mostly based on public resources, it also highlights some of the limitations of such activities. It was found that an entirely inclusive approach to volunteer contributions made the project subject to vandalism. [31] Additionally it illustrates the effect of relying on the availability of an individual for project maintenance: as of January 2012, all of his Pachube feeds have stopped updating. [32]

Safecast

The Safecast project offers an interesting contrast to the free-form activity of the Pachube community, and the solitary approach of Marian Steinbach. It was initiated by a well-connected team of experts, and benefits from collaborations with the Geiger counter manufacturer Medcom, and with Keio University in Tokyo who provides expertise in mapping and spatial analysis. Safecast is financially backed by a public fund raising effort and an undisclosed amount of private funding. All their work is executed by volunteers. [33][34][35]

The initial aim was to purchase radiation sensors and distribute them to volunteers in Japan who would be trained in their operations. Safecast now employs a mixture of such stationary sensors and ad hoc mobile measurements ("drives") by volunteers [36], using a range of radiation sensing devices, often augmented with custom hardware. Additionally they aggregate data from other public repositories, [37] and they offer the ability to contribute data in a public submission form. [38] Their data is in the public domain, and is presented in the form of high-quality maps. [39] Safecast chose to select contributors carefully, and to train them well. A blog post by a contributor illustrates the rigour involved in investigating a discrepancy with government-reported measurements. [40] Another blog post documents some of their training procedures. [41]

There are benefits to enhancing stationary sensor data with ad hoc "drives": it is possible to cover larger areas while relying on a limited supply of sensing equipment, it is possible to employ the short-term help of new volunteers (at the cost of training them) while retaining a consistent choice of sensing equipment, and their fairly rigorous procedures ensure a level of consistency. A downside is the lack of real-time updates for such measurements.

Safecast is a multi-layered organisation. It centralises the design of procedures and quality checks, which are in the hands of a few trusted individuals. It invites varying levels of contribution, subject to some social checks and balances, which opens up the project to a diverse set of potential contributors while offering some protection against vandalism. While their data is presented well it is hard to determine their data volume (the website quotes between 600,000 and 2,000,000 data points), and the degree to which sensing activity was sustained over time. In January 2012 they appear to have published data of about 30 new "drives". [36]

Summary

All collaborative efforts introduced here involved multiple data capturing approaches: data was often acquired in manual measurements, acquired from automated sensor stations, as well as aggregated from public sources. In all cases there was a strong focus on machine-readable data access via APIs and consistent data formats; a lot of the early efforts were purely focused on this aspect, usually by transcribing data found in unstructured, human-readable documents.

The general-purpose sensor data hub Pachube served as an important starting point for many efforts, both as a data platform, and as a meeting ground and collaboration channel for volunteers. Its infrastructure made it easy to publish sensor data, to demonstrate that collaboration efforts can lead to results, and that they can find a larger audience.

The Safecast project has demonstrated the benefits of checks and balances in collaboration efforts, the reliance on highly experienced specialists, and training procedures for volunteers. Their data reports are of a high quality and easy to navigate, and the overall project quality and level of collective expertise became a bargaining asset that lead to important collaborations with commercial and academic institutions.

These projects illustrate that there is demand for environmental monitoring systems that are accessible to an interested public, and that there is public motivation and ability to build such infrastructure where governments and other institutions fail to provide them. It is however evident that there still is limited experience in questions of organisational structure. One open question is how such efforts can be sustained while there is still public demand for them, particularly when virtually all activity is volunteer-driven.

References

[1] Denisa Kera, Connor Graham. Collective Sensor Networks and Future Communities: Designing Interaction across Multiple Scales. In *Proceedings of the 22nd*

Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction, 2010.

[2] http://www.wunderground.com/

[3] http://openenergymonitor.org/

[4] http://airqualityegg.wikispaces.com/

[5] Andrew Fisher. Towards a Sensor Commons, 2011. http://ajfisher.me/2011/12/20/towards-a-sensor-commons/

[6] M. Vinyals, J.A. Rodriguez-Aguilar, J. Cerquides. A Survey on Sensor Networks from a Multi-Agent perspective. In *2nd International Workshop on Agent Technology for Sensor Networks*, 2008.

[7] Cristina Gouveia, Alexandra Fonseca. New approaches to environmental monitoring: the use of ICT to explore volunteered geographic information. *Springer Science+Business Media B.V.*, 2008.

[8] Michael F. Goodchild. Citizens As Sensors: the World of Volunteered Geography, 2006.

[9] Jo Abbot, Robert Chambers, Christine Dunn, Trevor Harris, Emmanuel De, Gina Porter, Janet Townsend, Daniel Weiner. Participatory GIS: opportunity or oxymoron? In *PLA Notes 33, IIED*, 1998.

[10] Paul Schroeder. Criteria for the design of a GIS/2, 1996.

[11] Sacha Wunsch-Vincent, Graham Vickery. Participative Web - User-Created Content. *OECD Working Party on the Information Economy*, 2006.

[12] Jeffrey P. Cohn, Citizen Science: Can Volunteers Do Real Research? In *BioScience 58, AIBS*, 2008.

[13] Suvi Välimäki. A Case Study of Openstreetmap, *Master's Thesis in Geography and Geoinformatics*, 2011.

[14] Henry Mintzberg. The Structuring of Organizations: A Synthesis of Research. In *The theory of management policy series*, 1978.

[15] Marshall Scott Poole, Andrea B. Hollingshead, Joseph E. Mcgrath, Richard L. Moreland, John Rohrbaugh. Interdisciplinary Reasearch of Small Groups. In *Small Group Research 35*, 2004

[16] Mancur Olson. The Logic of Collective Action: Public Groups and the Theory of Groups. *Harvard University Press*, 1965.

[17] Elinor Ostrom. Governing the Commons: The Evolution of Institutions for Collective Action. *Cambridge University Press*, 1990.

[18] Mordechai Haklay, Sofia Basiouka, Vyron Antoniou, Aamer Ather. How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. In *The Cartographic Journal 47, Issue 4*, 2010.

[19] http://www.greenpeace.org/international/en/press/releases/Greenpeace-radiation-team-pinpoints-need-to-extend-Fukushima-evacuation-zone-especially-to-protect-pregnant-women-and-children-/

[20] http://blog.pachube.com/2011/03/real-time-radiation-monitoring-in-japan.html

[21] http://www.cbc.ca/spark/2011/06/full-interview-usman-haque-on-pachube/

[22] http://freaklabs.org/index.php/Blog/Misc/Hacking-a-Geiger-Counter.html

[23] http://blog.pachube.com/2011/03/calling-all-jp-users-distributed.html

[24] http://blog.pachube.com/2011/06/crowdsourcing-data-accuracy.html

[25] http://blog.pachube.com/2011/05/crowdsourcing-open-data-evolving-into.html

[26] http://blog.pachube.com/2011/03/crowd-sourced-realtime-radiation.html

[27] https://pachube.com/feeds?q=radiation&status=live

[28] http://www.sendung.de/2011-03-15/a-crowdsourced-japan-radiation-spreadsheet/

[29] http://www.sendung.de/2011-03-19/update-on-radiation-data-japan/

[30] http://www.cbsnews.com/8301-501465_162-20045784-501465.html

[31] http://www.sendung.de/japan-radiation-open-data/

[32] https://pachube.com/feeds?tag=bfs&status=live

[33] http://blog.safecast.org/about/

[34] http://spectrum.ieee.org/podcast/geek-life/hands-on/crowdsourcing-radiation-monitoring

[35] http://blog.safecast.org/history/

[36] http://maps.safecast.org/drives

[37] http://maps.safecast.org/feedmap

[38] http://maps.safecast.org/submit?lang=en

[39] http://blog.safecast.org/maps/

[40] http://blog.safecast.org/2011/06/discover-validate/

[41] http://blog.safecast.org/2011/10/a-high-school-student/