Fusing Human and Technical Sensor Data: Concepts and Challenges

Bernd Resch[1,2,3] and Thomas Blaschke[1]
[1]Interfaculty Department of Geoinformatics, University of Salzburg, Austria
[2]Heidelberg University, Germany
[3]Center for Geographic Analysis, Harvard University, USA

Abstract

As geo-sensor webs have not grown as quickly as expected, new, alternative data sources have to be found for near real-time analysis in areas like emergency management, environmental monitoring, public health, or urban planning. This paper assesses the ability of human sensors, i.e., user-generated observations in a wide range of social networks, the mobile phone network, or micro-blogs, to complement geo-sensor networks. We clearly delineate the concepts of People as Sensors, Collective Sensing and Citizen Science. Furthermore, we point out current challenges in fusing data from technical and human sensors, and sketch future research areas in this field.

1 Introduction

The predicted rise of geo-sensor webs has not taken place as rapidly as estimated approximately a decade ago. One may argue that this impedes a variety of research efforts from being carried out due to lacking near real-time base data. In contrast, we are currently witnessing the rapid emergence of user-generated data in a wide range of social networks, the mobile phone network, or micro-blogs. These human-generated data can potentially complement sensor measurements to a large degree, not by calibrated well interpretable measurements, but by subjective observations or human-generated measurements.

Current literature in the area of user-centred sensing oftentimes mixes up different approaches how data are generated, used and analysed. This paper distinguishes three concepts according to [10]. "1.) People as Sensors defines a measurement model, in which measurements are not only taken by calibrated hardware sensors, but in which also humans can contribute their subjective 'measurements' such as their individual sensations, current perceptions or personal observations. 2.) Collective Sensing tries not to exploit a single persons measurements and data, but analyses aggregated anonymised data coming from collective sources, such as Twitter, Flickr or the mobile phone network. 3.) Citizen Science stands for a human-based approach to science where citizens contribute semi-expert knowledge to specific research topics."

This paper discusses particular challenges in fusing data from human and technical sensors (s. Figure 1), comprising standardisation (on data, service and method levels), data assimilation (resolution, aggregation, etc.), multi-dimensionality in the data, combination of methods from geoinformatics and computational linguistics (to extract information from user-generated data), quality assurance (both for technical and human sensors), and the consideration of privacy issues (data ownership, storage, optimum aggregation levels, etc.), and last but not least the fusion of user-generated data with remote sensing data.

Figure 1: Increasing Availability of Technical and Human Sensors.

2 Concepts: People as Sensors, Collective Sensing and Citizen Science

Ubiquitous sensor networks can assist in decision-making in near real-time in a broad range of application areas such as public safety, traffic management, environmental monitoring or in public health [13]. Yet, analysing and monitoring our surroundings in near real-time is still a major challenge due to sparsely available data sources [10]. As a result from this shortcoming, coupled with the fast rise of mobile phones, a number of researchers have started to investigate alternative methods for generating real-time data relevant for decision-making processes. Recent efforts have been taken by OpenSignal [8], On Line Disaster Response Community [6], CenceMe [7] or Near Future Laboratory [4]. In scientific literature, we see a number of human-centred sensing approaches that can be summarised under three main concepts: People as Sensors, Collective Sensing and Citizen Science [10]. This sub-section presents a clear disambiguation between these concepts.

"People as Sensors" defines a sensing model, in which measurements are not only taken by calibrated hardware sensors, but in which also humans can contribute their subjective measurements such as their individual sensations, current perceptions or personal observations [12]. Like this, people act as non-technical sensors with contextual intelligence and comprehensive knowledge. Measurements are not created absolutely reproducibly by calibrated sensors, but through personal and subjective observations. Such observations could be air quality impressions, street damages, weather observations, or statements on public safety, submitted via dedicated mobile or web applications. A vibrant real-world example is WAZE [14], a smartphone app allowing people to send their personal traffic reports, which are directly used in other persons routing requests. These human sensors can thus complement–or in some cases even replace–specialised and expensive sensor networks. Throughout recent literature, the term "People as Sensors" is used interchangeably with "Citizens as Sensors" [5] or "Humans as Sensors" [3].

A concept related to People as Sensors is Participatory Sensing, in which a number of persons with a common goal in a geographically limited area contribute geo-referenced data via their end user devices such as smartphones [2]. From this definition it is evident that the term Participatory Sensing is highly similar to People as Sensors, but its definition is a little more restricted in terms of input devices, data acquisition and information processing.

Furthermore, we are currently witnessing a fast rise of **Collective Sensing** approaches. In contrast to People as Sensors, Collective Sensing is an infrastructure-based approach, which tries to leverage existing Information and Communications Technology (ICT) networks to generate contextual information. This methodology tries not to exploit a single persons measurements and data. Collective Sensing analyses aggregated and anonymised data coming from collective networks, such as Flickr, Twitter, Foursquare or the mobile phone network. Like this, we can gain a coarse picture of the situation in our environment without involving personal data of individual persons.

Finally, the term **Citizen Science** plays a key role in the context of People as Sensors. Citizen Science basically states that "through the use of sensors paired with personal mobile phones, everyday people are invited to participate in collecting and sharing measurements of their everyday environment that matter to them" [9]. An example for promoting the Citizen Science concept is the Citizen Science–Community Involvement Today and in the Future grant program by the US Environmental Protection Agency (EPA). This program aims to encourage individuals and community groups in New York City to collect information on air and water pollution in their communities, and seek solutions to environmental and public health problems.

Table 2 [10] summarises the comparison of the discussed concepts of People as Sensors, Collective Sensing and Citizen Science according to the following criteria.

- *Voluntary/Involuntary*: whether contributing people voluntarily (dedicatedly) share their data for further (geo-spatial) analysis or decision-making
- *Content*: type of data, which are contributed
- *A Priori Knowledge*: required knowledge of the user
- *Contextual Data*: whether the contributed data contain contextual intelligence, for instance a person's local knowledge
- *Reliability*: quality level of the generated data and contributors' trustworthiness
- *Analysed Datasets*: whether single (individual) datasets are analyzed or spatially and temporally aggregated (anonymised) data are used
- *Specific Infrastructure*: whether additional dedicated infrastructure is necessary to collect data

	People as Sensors	Collective Sensing	Citizen Science
Voluntary/Involuntary	Voluntary	Involuntary	Voluntary
Content	Layman Observations	Raw geo-data (images, tags, ...)	Semi-professional Observations
A Priori Knowledge	Medium	Low/None	High
Contextual Data	Yes	Yes	Yes
Reliability	Medium	Mediocre	Good
Analysed Datasets	Individual	Aggregated	Individual
Specific Infrastructure	No	No	No

Table 1: Comparison of Human-Centric Sensing Concepts[10]

3 Challenges in Fusing Human and Technical Sensor Data

In bringing the measurement concepts laid out in section 2 to practical use for data analysis, visualisation and communication for decision support, the central issue is how to fuse data from technical and human sensors coming from a variety of sources as illustrated in Figure 2. This section illustrates the most pressing challenges, ranging from standardisation, quality assurance and fusion of different sampling rates to the representativeness/reliability of user-generated data and data privacy and protection issues.

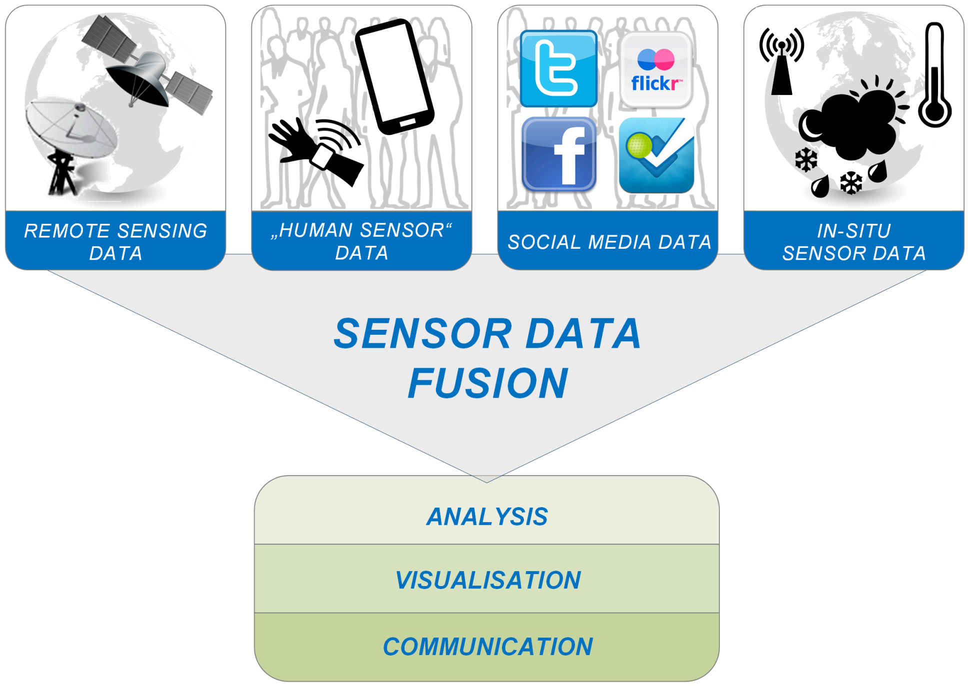

Figure 2: Increasing Availability of Technical and Human Sensors.

The first challenge in fusing human and technical sensor data is the pertaining lack of standardisation. **Standardisation on data, service and method** levels fosters interoperability as a basis for joint analyses of human and technical sensor data. Even though reasonably mature data and service standards exist for technical sensors, for instance, through the OGC Sensor Web Enablement initiative [1], the integration of human sensor data including measurement processes, uncertainties, context variables, etc. is not possible yet.

A further challenge is the fusion of data from sensors with **diverse sampling rates and spatial resolutions**. First, this applies to sensors measuring in different temporal intervals (e.g., one sensor measuring in 3s intervals and the other one measuring in 7s intervals). Intelligent methods have to be found to combining these data apart from the least common multiple, which could be achieved by interpolation algorithms based on dynamic regression models depending on the sensors' characteristics. Another dimension of complexity is added when attempting to fuse in-situ sensor data with remote sensing data regarding temporal measurement intervals, which induces additional challenges through the combination of point data with raster data.

Another challenge that is particularly related to analysing data from social media channels. Here, most previous geo-analysis methods focused on finding spatial patterns, neglecting the **multi-dimensional nature of social media data**. Thus, new methods need to be developed across the disciplines of GIScience, temporal analysis, computational linguistics, computer science, and others to fully leverage the potential of social media data. This also includes the correlation of social media data with "People as Sensors" observations and technical data to maximise information density. One example of such an integrated approach is demonstrated in [11].

A highly complex, but overly important issue is **quality assurance in technical and human sensor data**. For technical sensors, a number of methods exist (Kalman filtering, threshold detection, dynamic tolerance levels, comparison with neighbouring stations, flatline checks, spatial regression or comparison with modelled data) [15] as the properties of the measurement procedure are known. However for human sensor data, traditional geo-data quality parameters as defined by the International Organization for Standardization (ISO) including accuracy, position accuracy, completeness, consistency, lineage, etc. need to be extended by new parameters such as expertise of the human sensor, spatial and temporal plausibility, or up-to-dateness. Even though these parameters might be rather simple to be defined, their actual assessment is highly difficult due to the unknown measurement process characteristics in human sensors.

The different nature of user-generated data results in differences in terms of **semantic expressiveness of mobile phone data**: The user-generated mobile network traffic represents a relatively large proportion of the population across social classes. However, these data are typically lacking content. For instance, the number of text messages sent or received might be logged, rather than the text itself; or the number and duration of voice calls might be logged, rather than the topic of the talk itself. This is in strong contrast to social media data and VGI, which are typically generated by a rather specific sub-group of the population, and explicitly contain content of some semantic value.

A connected methodological issue in the field of semantics is **representativeness in VGI**. This has to be tackled by a combined bottom-up/top-down approach. In bottom-up approaches, user groups and communities define their own semantic objects and interrelations between these in separate taxonomies. In contrast, top-down approaches try to define semantic rules and ontological relations as generically as possible–mostly before actual applications exist and decoupled from real-world use cases. Only the combination of those approaches can result in trans-domain semantic models, which are linked via object relations.

The requirement of high-quality information seems to be self-evident, but has not been tackled thoroughly for real-time geo-sensor networks and People as Sensors based approaches. The subjectivity of human "measurements" naturally raises the question of **trustworthiness** of these data in terms of data quality. As discussed above, this results in uncertainty in the observation data. Thus, automated quality assurance mechanisms need to be developed for uncertainty estimation, dynamic error detection, correction and prevention. Different approaches are in development, e.g., Complex Event Processing (CEP) for error detection, standardisation efforts for representing uncertainty in sensor data (e.g. Uncertainty Markup Language–UncertML [16], or proprietary profiles to define validity ranges for particular observations. Such issues need to be solved in order to ensure reliability of both technical and human sensor data.

Another pressing question is: how can we preserve people's **privacy** when dealing with user-generated data and information, and partly sensitive personal data, in the context of mobile phones as ubiquitous in situ geo-sensors? In terms of privacy, the claim might arise that we need to be aware of our personal and private data before we share them. This also raises the need to discuss the concept of U-VGI, i.e. Un-Volunteered Geographic Information, in contrast to VGI [10]. For instance, Collective Sensing approaches exploit anonymised data from digital networks (e.g. by deducing crowd movements from traffic distribution in the cell phone network) even though people have not intended to share their data in this way.

As mobile phone data and human sensor data are individual oftentimes sensitive, **legal frameworks** have to be developed on national, trans-national and global levels to protect those personal data. The largest limiting factor in this regard is the varying interpretation of 'privacy' in different parts of the world. For instance, privacy can be traded like an economic good by its owner in the USA, whereas it is protected by law in the European Union. This means that supra-national legislation bodies and initiatives are called upon to set up appropriate world-wide regulations.

This also includes the critical question of **data ownership**. Shall they be owned by data producers, i.e., the citizens or a mobile phone network operator? Or rather the institution that hosts a system in order to collect data? Or the data providers? Furthermore, if sensitive data is analysed to produce anonymised information layers, who is responsible if decisions that are based on this information are wrong due to lacking quality of the base data? In conclusion, the issues of privacy, data ownership, accessibility, integrity and liability have to be tackled thoroughly all at once and not separately from each other.

Finally, **remote sensing** (RS) data are a somewhat special case within the realm of human and technical sensor data fusion. RS data are typically well structured-and may therefore not fall under the 'big data paradigm' even though such data sets can exceed the Terrabyte dimension. RS may often provide data of good comparability and repeatability (i.e., repeated measures for the same variables) and clearly defined types and levels of resolution. Until recently, the level of resolution of remotely sensed data was often seen critical for the integration with in-situ data, particularly the types of data described herein. Different societal research questions require different levels of resolution in space and time, and perhaps also regarding the spectral resolution. While 2015 marks a new dimension in spatial resolutions of civilian, publicly accessible satellite remote sensing–WorldView-3 offers 30 cm panchromatic and 1,24m multispectral resolutions – the variety of needs in human sensing suggests that any data fusion methodology should be highly flexible. We may conclude that since for RS data all meta-data are plannable and known the gained evidence about the world is predominantly dependent on the point and time of the observation. We may epistemologically call this realism–some may call it positivism while assuming that objects exist independently from the observer.

4 Conclusion

Without any doubt, fusing human and technical sensor data includes a wide variety of technical and methodological challenges. It may, however, comprise major epistemological problems. When combing data from different sources gathered under different schemas and for different purposes, one faces the potential problems of fusing 'apples and pears'. This short article widely follows a techno-positivist epistemology while assuming that an objective reality can be explored and determined through scientific methods–basically observation and testing which create verifiable knowledge about the world. Obviously, the limits to this approach are not only determined by our ability to measure precisely. The key question is if we are measuring the right, i.e., relevant variables to the underlying societal questions.

This article categorises some of the major challenges in this predominantly technical domain. The exception to the technical dominance of this field and the potential lynchpin for supporting societal relevant research could be the citizen science approach. However, being a young development many citizen science projects may be at risk to be regarded as an end in itself: it is often investigated whether or not citizen science works and examples where citizen science was or is the key to solve a real-world problem which could not be solved otherwise are rare.

A scientific challenge lies in the trivial question whether evidence about the real world depends upon the perspective of the observer: Two persons who view the same object may interpret it quite differently because of their different personal history and assumptions about the real world. While GIS systems are technically mature and so may sensor devices be, we believe that the two major concepts discussed herein, namely 'people as sensors' and 'collective sensing'–in combination with citizen science methods–are still in their infancy and need methodological and epistemological foundations.

References

[1] M. Botts, G. Percivall, C. Reed, and J. Davidson. Ogc® sensor web enablement: Overview and high level architecture. In *GeoSensor networks*, pages 175–190. Springer, 2008.

[2] J. A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. *Center for Embedded Network Sensing*, 2006.

[3] B. Forrest. Humans as sensors. *LBX Journal*, 21, 2010.

[4] F. Girardin, F. Calabrese, F. D. Fiore, C. Ratti, and J. Blat. Digital footprinting: Uncovering tourists with user-generated content. *Pervasive Computing, IEEE*, 7(4):36–43, 2008.

[5] M. F. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007.

[6] M. Laituri and K. Kodrich. On line disaster response community: People as sensors of high magnitude disasters using internet gis. *Sensors*, 8(5):3037–3055, 2008.

[7] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell. Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 337–350. ACM, 2008.

[8] A. Overeem, J. R Robinson, H. Leijnse, G.-J. Steeneveld, B. P Horn, and R. Uijlenhoet. Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophysical Research Letters*, 40(15):4081–4085, 2013.

[9] E. Paulos. Citizen science: Enabling participatory urbanism. In M. Forth, editor, *Handbook of Research on Urban Informatics: The Practise and Promise of the Real-time City*, Lecture Notes in Geoinformation and Cartography, pages 414–436. Springer Berlin Heidelberg, 2009.

[10] B. Resch. People as sensors and collective sensing-contextual observations complementing geo-sensor network measurements. In J. M. Krisp, editor, *Progress in Location-Based Services*, Lecture Notes in Geoinformation and Cartography, pages 391–406. Springer Berlin Heidelberg, 2013.

[11] B. Resch, A. Summa, G. Sagl, P. Zeile, and J.-P. Exner. Urban emotionsgeo-semantic emotion extraction from technical sensors, human sensors and crowdsourced data. In *Progress in Location-Based Services 2014*, Lecture Notes in Geoinformation and Cartography, pages 199–212. Springer Berlin Heidelberg, 2015.

[12] G. Sagl, B. Resch, and T. Blaschke. Contextual sensing: Integrating contextual information with human and technical geo-sensor information for smart cities. *Sensors*, pending, 2015.

[13] S. Sneha and U. Varshney. Enabling ubiquitous patient monitoring: Model, decision protocols, opportunities and challenges. *Decision Support Systems*, 46(3):606–619, 2009.

[14] Waze. Free gps navigation with turn by turn, 2015.

[15] A. Wieser. Smarte qualitätssicherung von hydrologischen messdaten in naher echtzeit. *Muenchner Fortbildungsseminar, Geoinformationssysteme*, pages 279–285, 2013.

[16] M. Williams. Uncertainty markup language (uncertml). *Technical Report OGC Discussion Paper 08-122r2, Version 0.6*, 2009.