Capturing the Long Tail of Sensor Web

Steve H.L. Liang, David Chang, James Badger, Rohana Rezel,
Shawn Chen, Chih-Yuan Huang, Ren-Yu Li

Department of Geomatics Engineering, University of Calgary
2500 University Drive NW, Calgary, Alberta, Canada
{steve.liang, ycchang, jpbadger, rdrezel, hschen, huangcy, ryli}@ucalgary.ca

Keywords: Volunteered Geospatial Information, Sensor Web, Web 2.0, Cyberinfrastructure, Virtual Globe

1. Sensor Web

Large-scale sensor networks and the vast data sets they produce worldwide are being utilized and published by a rising number of organizations on an ever-increasing frequency. Examples include the global scale ARGOS network of buoys[1], the weather networks of the World Meteorological Organization, the global GPS Zenith Total Delay (ZTD) observation network, etc. Significant amount of efforts (e.g., GEOSS[2] and NOAA IOOS[3]) have been put forth to web-enable these large-scale sensor networks so that these sensors and their data can be accessible through interoperable sensor web standards. Moreover, with the advent of the low-cost sensor networks and data loggers, it is technologically and economically feasible for individual scientists to deploy and operate small to medium scale sensor arrays at strategic locations for their own research purposes. There is a spectrum of sensor networks ranging from local scale short-term sensor arrays to global scale permanent observatories. The vision of a worldwide sensor web is becoming a reality.

2. Long Tail

Chris Anderson [1] popularized the concept of long tail in his book *The Long Tail: Why the Future of Business is Selling Less of More*, in which he mentioned Netflix as examples of business applying this strategy in movie rental business. For example, before the movie rental web sites, such as Netflix, customers can only rent movies from physical movie rental stores, such as Blockbuster™. However, physical stores concentrate on offering popular movie titles, *i.e.*, focusing only on the head of the long tail. Less popular movies are not available in these physical movie rental stores. Customers may not even know of the existence of movies that they might be interested because they cannot see them on the shelves in their local movie rental stores.

Long tail has been found in many different businesses. Similarly, we find long tail phenomenon repeats itself in sensor web. Today most researchers rely on the sensor data portals hosted by few large-scale government and research institutes (e.g., GEOSS, NOAA, NASA, Environment Canada) to search for and download sensor measurement data. The number of these portals is relatively small, but these portals provide substantial amount of continuous and long-term sensor measurements. These sensor data are quality-controlled, calibrated, indexed, stored, and published through the web. We can call these sensors and their data the "head" in the sensor web long tail graph (Figure 1), and the head of sensor web is currently accessible to the researchers.

However, the number of local scale sensor networks, deployed and operated by individual scientists, is growing exponentially in recent years. These sensors/data are not accessible to most of the researchers. Heidorn [2] in his article *Shedding Light on the Dark Data in the Long Tail of Science* described this particular troublesome class of data "Dark Data". They are not indexed nor stored at the centralized sensor data portal, so they become nearly invisible to most scientists and other potential users. These sensors/data are more difficult to find and less frequently reused. As a result, they are more likely to remain underutilized and eventually lost. These sensors and their measurements comprise the "tail" in the sensor web long tail graph (Figure 1).

3. Challenges in Capturing the Sensor Web Long Tail

Building such system to capture the sensor web long tail needs to overcome the following challenges.

(1) Handling the heterogeneity: The head of sensor web consists of few organizations generating huge amount of datasets. Therefore, they tend to be much more homogeneous, and this uniformity makes the computing and

[1] http://www.argos-system.org
[2] http://www.earthobservations.org/geoss.shtml
[3] http://ioos.gov/

data management issues simpler. On the other hand, the tail consists of large number of individual sensor publishers. As a result, the tail is much more heterogeneous (e.g., proprietary data formats or different sensor hardware). How to design a system allowing users to access the tail's heterogeneous sensor resources in a coherent and homogeneous manner is the first major challenge; *(2) protecting data ownership:* It's common for researchers to spend more than 50% of their time and funding on data collection. Without proper mechanisms to protect sensor/data ownership, scientists will not publish and share their sensors and data;

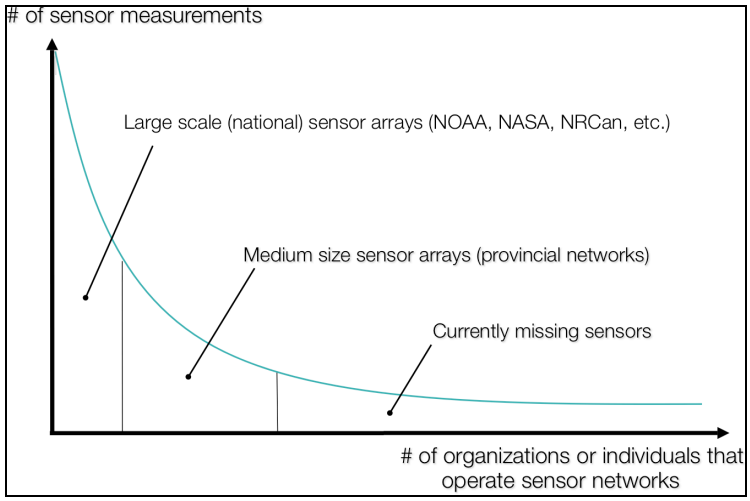

Figure 1 Sensor Web Long Tail Graph

(3) motivating scientists to publish sensors/data: There should be incentives to motivate scientists to contribute their sensors/data to the system. The system should be able to create a network effect such that the more high quality sensors/data an individual contribute, more and more value the individual will gain, and, as a consequence of the network effect, more sensor owners will be attracted to publish and share their sensing resources; *(4) creating a rich user experience:* The number of web content publishers as well as the amount of web content grew exponentially when the intuitive and easy-to-use web publishing tools became available (e.g., blogs, wikipedia, and twitter). Similarly, in order to capture the sensor web long tail, the system should provide an intuitive and coherent user interface allowing users to search for, browse, preview, download and publish sensors/data.

4. Capturing the Sensor Web Long Tail

At the GeoSensor Web Lab at the University of Calgary, we are designing an architecture and building a crowd sourcing-based sensor web platform called GeoCENS (Geospatial Cyberinfrastructure for Environmental Sensing)(See Figure 2). Our aim is to propose innovative approaches and provide the missing software components for capturing the currently missing sensor web long tail. GeoCENS proposes and implements the following solutions in order to address the aforementioned challenges: (1) we designed and implemented an Online Social Network-based (OSN-based) sensor web architecture that exploits sensor web users' collective intelligence. For example, its underlying social graphs, the structure of user interactions, and the users' profiles/preferences allows us to develop a sensor web search engine based on the modern web search algorithms (e.g., Google's PageRank [3]); (2) Recommendation engines are key components in existing non-geospatial long tail systems such as Pandora[4] and Rhapsody[5]. With the GeoCENS social network infrastructure, we are able to design and develop a sensor web recommendation engine [4] (*i.e.*, a geospatial folksonomy and a collaborative tagging system) that recommends sensors and datasets according to a user's geographical area of interest; (3) we uses the OGC sensor web framework [5] to handle the sensor web heterogeneity and facilitate sensor web interoperability; (4) we are employing digital watermarking algorithms in order to protect the sensor data ownership; and (5) we develop an intuitive 3D sensor web browser allowing users to maneuver a 3D virtual globe, in order to browse, visualize, access and publish heterogeneous sensing resources and other relevant information. To our best knowledge, GeoCENS is the first OSN-based sensor web platform, the first sensor web recommendation engine as well as the first virtual globe-based sensor web browser.

[4] http://www.pandora.com
[5] http://www.rhapsody.com

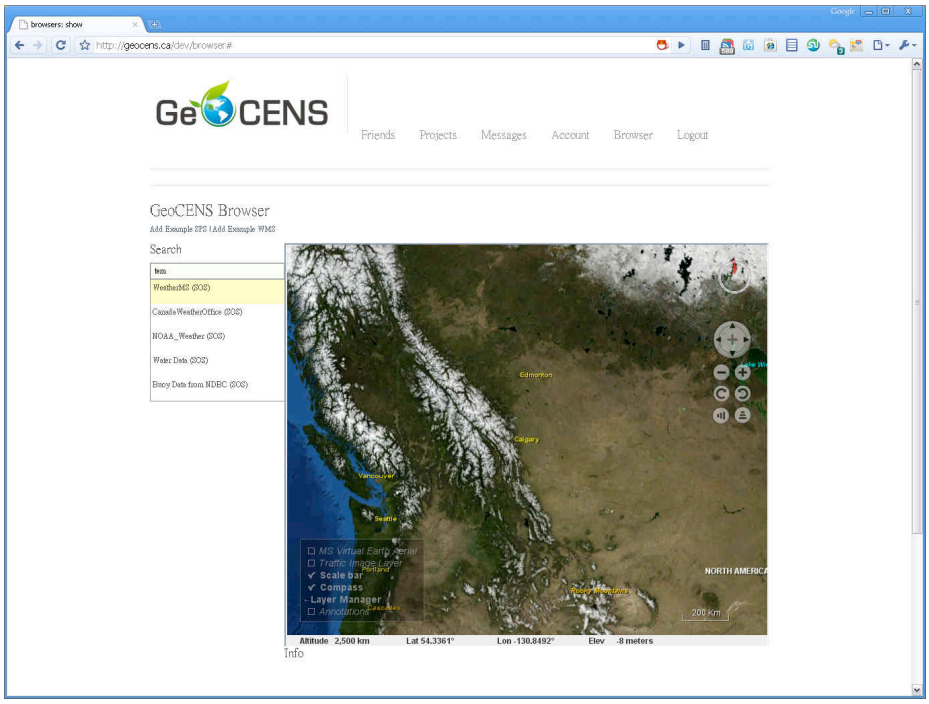

Figure 2 A screen capture of the GeoCENS platform. Left hand side shows a sensor web search engine, and a list of suggested sensor web resources (according to the user entry keyword as well as the user's current viewport).

References

[1] C. Anderson, *The Long Tail: Why the Future of Business is Selling Less of More*, Hyperion, 2006.

[2] P.B. Heidorn, "Shedding light on the dark data in the long tail of science," Library Trends, vol. 57, 2008, pp. 280–289.

[3] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," Computer Networks and ISDN Systems, vol. 30, Apr. 1998, pp. 107-117.

[4] R. Rezel and S.H. Liang, "SWE-FE: Extending Folksonomies to the Sensor Web," Proceedings of the Workshop on Sensor Web Enablement 2010 (SWE2010), Chicago, IL: 2010.

[5] M. Botts, G. Percivall, C. Reed, and J. Davidson, "OGC® sensor web enablement: Overview and high level architecture," GeoSensor Networks, 2008, pp. 175–190.