

GeoCENS: Geospatial Cyberinfrastructure for Environmental Sensing

S.H.L. Liang¹, D. Chang¹, J. Badger¹, R. Rezel¹,
S. Chen¹, C.Y. Huang¹, R.Y. Li¹

¹University of Calgary, 2500 University Dr. NW, Calgary, AB, CANADA
Email: {steve.liang, ycchang, jpbadger, rdrezel, hschen, huangcy, ryli}@ucalgary.ca

1. Introduction

In recent years, large-scale sensor arrays and the vast data sets they produce worldwide are being utilized, shared and published by a rising number of researchers on an ever-increasing frequency. With the rapidly increasing number of large-scale sensor network deployments, the vision of a World-Wide Sensor Web (WSW) is becoming a reality [1]. Similar to the World-Wide Web (WWW), which acts essentially as a “World-Wide Computer”, the Sensor Web can be considered as a “World-Wide Sensor” or a “cyberinfrastructure” that instruments and monitors the physical world at temporal and spatial scales that are currently impossible. However, realizing the worldwide sensor web vision is very challenging. Building a sensor web system requires addressing the following open problems:

- (1) Discovering relevant data among the distributed sensors and delivering it to interested users efficiently
- (2) Handling heterogeneous sensor networks and their data independently of the underlying network protocols, hardware, data models, and data formats.
- (3) Preventing transfer large volumes of sensor data streams across the network
- (4) Handling large numbers of sensors, and large numbers of users.

In the GeoCENS (Geospatial Cyberinfrastructure for Environmental Sensing) project, we are designing an architecture and building a sensor web platform. With GeoCENS, users can maneuver a 3D sensor web browser, within a single virtual globe, in order to discover, visualize, access and share heterogeneous and ubiquitous sensing resources, and other relevant information. Our aim is to address the aforementioned technical challenges, propose innovative approaches, and provide the missing software components for realizing a worldwide sensor web.

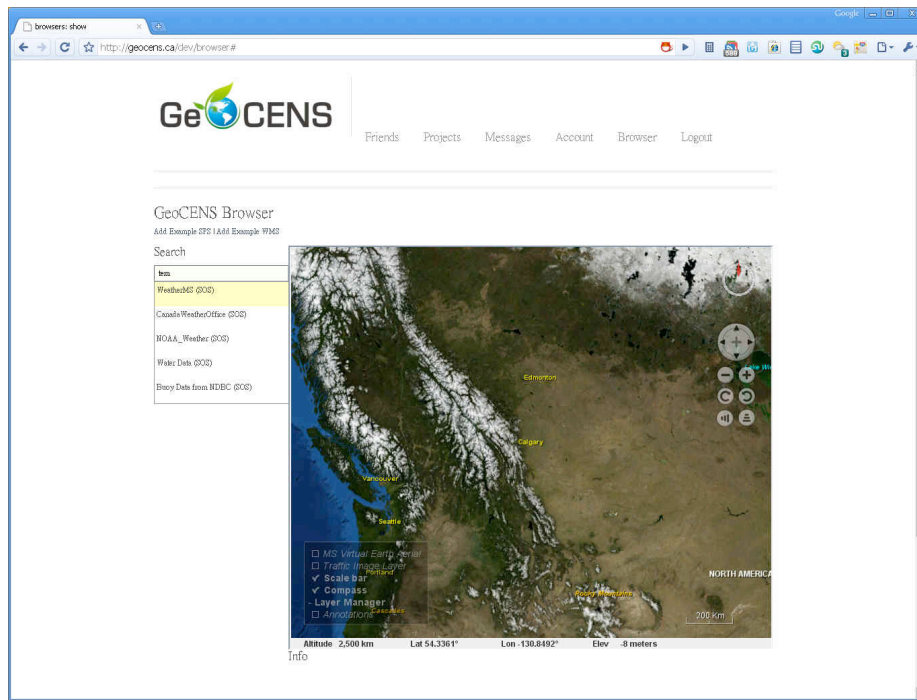


Figure 1. A screen capture of the GeoCENS platform. Left hand side shows a sensor web search bar, and a list of suggested sensor web resources (according to the user entry keyword as well as the user's current viewport).

2. GeoCENS Architecture

The key features of the GeoCENS architecture are highlighted in this section. Detailed architecture as well as a live demonstration will be presented in the paper presentation.

2.1 Online Social Network (OSN)

GeoCENS is an OSN-based sensor web platform for researchers. Take facebook.com, today's most popular OSN [2], as an example. On facebook users can share photos/activities with their friends and networks. On GeoCENS researchers interact and share sensors, scientific datasets (including data from sensors), experiences, and activities with their friends (e.g., colleagues from other institutes) and networks. Each GeoCENS user creates a profile where he/she declares his/her research interests and preferences, and establishes friendship with other users. A "*friendship*" is formed on GeoCENS when one GeoCENS user extends an (friendship) invitation to another user. Upon confirmation by the latter, the friendship relationship is formed. Other features include the ability to upload sensor datasets, the ability to join projects/groups of shared (geographical) area of research interest, the ability to adjust different privacy levels, and the ability to review/annotate/rate sensors as well as datasets.

By creating a specialized OSN for sensor web users, our goal is to leverage the underlying social graphs, the structure of user interactions, and the users' profiles/preferences to create innovative uses and applications of the sensor web. One innovative OSN-based sensor web application is a sensor web recommendation engine. We will describe the sensor web recommendation engine in the next section. We have considered building GeoCENS on top of an existing open source OSN framework. We investigated and compared three systems, namely Ning.com, Drupal.org, and Elgg.org. However, each system has its limitations, and we decided to

build GeoCENS from scratch. We chose Ruby-on-Rail as our web development framework.

2.2 Sensor Web Recommendation Engine

With the GeoCENS social network infrastructure, we are able to develop a sensor web recommendation engine (*i.e.*, a collaborative tagging system) that recommends sensors and datasets according to a user's geographical area of interest. In fact, existing folksonomy-related research is mostly focused on non-geospatial applications [3]. One key contribution of the GeoCENS recommendation engine is that we extend the folksonomy research into geospatial applications by leveraging the geospatial information associated with three key components of collaborative tagging systems: *tags*, *resources* and *users*.

GeoCENS recommendation engine provides the following three functions: **(1) Collaborative tagging:** It enables the user to assign tags to the resources (e.g., assign multiple tags to a sensor). In order to make the task easier for the user and to avoid ambiguity, our recommendation engine is able to suggest tags to the user. We designed a new algorithm to suggest tags taking geospatial characteristics of the sensor web resources into consideration. **(2) Collaborative browsing:** This function enables users to navigate through the tags collected in the system. It aids in the process of sensor and sensor data discovery. We developed a new algorithm for building tag maps, a tag cloud for a sensor web browser where the geospatial attributes of the tag assignments are taken into consideration. Figure 2 shows a screen capture of an example GeoCENS tag map. **(3) Collaborative searching:** This function enables users to retrieve resources based on tag queries, either by clicking through a tag cloud, or by typing them out. The key is to retrieve the most relevant results for these queries, and we proposed an algorithm to enhance the query processing by taking geospatial aspects of the queries and data into consideration.

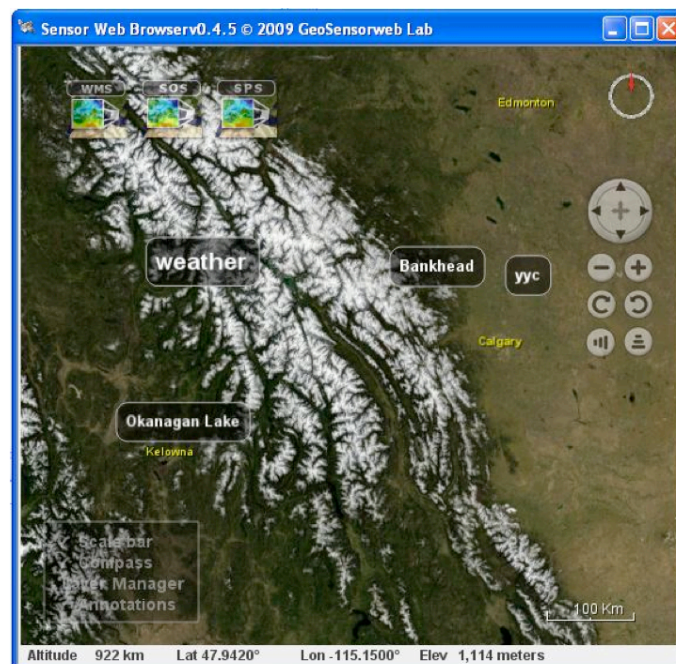


Figure 2. A screen capture of GeoCENS Recommendation Engine's tag map. A user can mouse click on a tag, and the system will load the corresponding tag assignments (*i.e.*, a list of OGC SOS services) from the recommendation engine server. Then the

user can choose his/her interested SOS service, and the browser will load sensor observations from the selected SOS service and display them on the map.

2.3 Decentralized Hybrid P2P Sensor Web Service Discovery

For any large-scale distributed system, both communication and data management distill down to the problem of resource discovery. Similarly, GeoCENS needs a sensor web resource discovery service. In order to handle sensor web's large numbers of sensors and large numbers of users, GeoCENS uses a hybrid P2P architecture for sensor web resource discovery. Every GeoCENS sensor web server (e.g., every OGC sensor web server) also serves as part of the sensor web service discovery infrastructure. These nodes operate on a cooperative model, where each peer leverages each other's available resources (*i.e.*, CPU, storage, bandwidth, etc.) for mutual benefit.

From literature and existing systems, there are two types of P2P architectures: (1) unstructured P2P networks [4] and (2) structured P2P networks [5][6]. Un-structured P2P networks are networks where participating nodes perform actions for each other, where no rules exist to define or constrain connectivity between nodes. They are simple but not scalable because their flood-based query processing generates enormous amounts of network traffic. Structured P2P networks use hash functions to build distributed indexes for their stored data items. The hash tables, like distributed indexes, successfully reduce the nodes to be scanned per query. However, structured P2P networks are vulnerable to node dynamic.

GeoCENS' hybrid approach means we are using both structured and unstructured P2P networks. The rationale of such hybrid design is described as follows. We envision that the future sensor web will have two types of sensor web servers: (1) Powerful sensor web servers maintained by large institutions (e.g., NASA or NOAA). They would not join and leave the network randomly; in most cases these servers are made accessible 24-7, that means they are static nodes in the network; and (2) Less powerful sensor web servers maintained by small institutions or even individuals (e.g., Universities or citizen scientists). These servers might join and leave the network more frequently, that means they are dynamic and transient nodes in the network. Considering the above-described settings it is a rationale design decision to group static P2P nodes into structured super-nodes (to exploit the stability of static nodes) and group dynamic P2P nodes into leaf-nodes (to save the overhead for maintaining the structure). Since structured P2P networks can only process exact key-value pair queries, we enable geospatial search functions by labeling data with Space Filling Curves (SFC). Our architecture is also unique in that it is a locality-aware system, *i.e.*, the system is able to exploit the locality information between peer nodes in order to deliver the query results quickly and efficiently.

2.4 OGC-based Sensor Web Services

GeoCENS uses the Open Geospatial Consortium (OGC) Sensor Web Enablement (SWE) architecture as the fundamental interoperability architecture. SWE is an OGC initiative that is building an open and interoperable Geospatial web service framework to enable the exchange and processing of sensor observations and relevant sensor web data [7]. GeoCENS implemented the OGC Sensor Observation Service specification (SOS), SensorML specification, and Observation and Measurement specification (O&M) [8]. GeoCENS SOS implementation is unique in that we optimize the performance for virtual globe systems. We developed a parallel processing framework to pre-process raw sensor readings. MapReduce tasks are performed in order to index

and aggregate the raw readings into multiple “tiles” at different resolutions. We use Java and CouchDB to implement the GeoCENS server. Initial benchmark results demonstrate that our server implementation outperforms the open-source 52North¹ SOS implementation.

2.5 Virtual Globe-based Sensor Web Browser

The GeoCENS browser is an intuitive 3D client frontend for all GeoCENS services. It allows users to maneuver a 3D sensor web browser, within a single virtual globe, in order to browse, discover, visualize, access, share and tag heterogeneous sensing resources and other relevant information. Starting from a ‘zoomed out’ view of the globe, users are able to select a study site and ‘fly’ into it. While flying to their study sites, multiple resolution *in-situ* sensing data are loaded to the client from the distributed GeoCENS SOS servers. The GeoCENS browser combines multiple sensor data streams and geographical datasets, and render them in a coherent and unified virtual globe environment.

We develop the GeoCENS browser on top of the open source WorldWind virtual globe system². To our best knowledge, it is the world’s first OGC-based sensor web 3D browser. The GeoCENS browser has the following three unique components/contributions: (1) In order to interoperate with existing sensor web servers, we developed an OGC SWE communication module that is able to communicate with OGC SWE-compatible servers; (2) In order to reduce unnecessary data transfer in the 3D virtual globe environment, we developed a new sensor data loading module based on Hierarchical Triangular Meshes (HTM) [11] instead of using WorldWind’s Quadtree tile based loading module; and (3) In order to prevent transferring large volume of sensor data across the network repeatedly, we developed a client-side cache that indexes, aggregates, and caches spatial-temporal sensor data.

3. Related Work

Several recent works have attempted to propose architectures for sensor web systems. Intel Research’s IrisNet [10] proposes a decentralized architecture based on a hierarchical topology and provides techniques to process queries over a distributed XML document containing sensor data. However, IrisNet only supports very preliminary geospatial queries. It uses hierarchical place names to build its hierarchical network topology. In order to perform geospatial query, users/applications need to know the exact place name *a priori* and explicitly specify the parts of a hierarchy that the query needs to traverse. Moreover, IrisNet does not have a sensor discovery module and can only handle homogeneous sensors and datatypes.

Microsoft Research’s SensorMap [11] uses a centralized web portal design, and tackles the scalability and performance issues by building a COLR-Tree, a data structure that indexing, aggregating and caching sensor streams in order to prevent transferring large volume of sensor streams across the network. However, SensorMap’s centralized design makes the portal a single point of failure. In contrast, GeoCENS uses a Service-Oriented Architecture (SOA) and a decentralized hybrid P2P architecture for sensor service discovery. There is no single point of failure, and the P2P sensor discovery service balances the load by directing queries and traffic to the distributed sensor web services. Moreover, both IrisNet and SensorMap use its own proprietary interfaces and sensor data encodings while GeoCENS follows OGC Sensor

¹ <http://www.52north.org>

² <http://worldwind.arc.nasa.gov/>

Web Enablement (SWE) specifications [7] and is able to interoperate with other OGC-compliant sensor web servers.

More importantly, GeoCENS is innovative and unique in that it is a social network-based sensor web platform. GeoCENS harvests the sensor web users' interaction structures and activities in order to build innovative sensor web applications. For example, with its social network infrastructure GeoCENS is able to build a geospatial folksonomy for the sensor web that recommends relevant sensor web resources to a user according to the collective intelligence of the GeoCENS users.



Figure 3. A screen capture of the GeoCENS sensor web browser. The icons are the sensor locations from the selected OGC SOS server. The pop-up window shows the latest sensor observations of that particular sensor. In this case, the pop-up window shows two independent sensor observations. These two observations observed the same feature of interest (FOI), *i.e.*, YQL airport. These two observations observed two different phenomena: relative humidity and temperature. Their values are 87% and -0.2 Celcius respectively, and these two observations were collected at the same timestamp (*i.e.*, 2010-02-15T19:42:00-0700).

Acknowledgements

GeoCENS project is sponsored by Canarie's NEP-2 funding programme, Cybera Inc., and NSERC Discovery grant.

References

- Balazinska M et al., 2007, Data Management in the Worldwide Sensor Web. *IEEE Pervasive Computing*, 6: 30–40.
- Nazir A, Raza S, and Chuah C, 2008, Unveiling Facebook: A Measurement Study of Social Network Based Applications. *IMC'08*. Vouliagmeni, Greece.
- Hotho A, Jaschke R, Schmitz C, and Stumme G, 2006, Information retrieval in folksonomies: Search and ranking. *Lecture Notes in Computer Science*, vol. 4011:411.
- Ripeanu M, 2001, Peer-to-Peer Architecture Case Study: Gnutella Network, *First International Conference on Peer-to-Peer Computing (P2P'01)*, Linkoping, Sweden.

- Rowstron A and Druschel P, 2001, Pastry: Scalable, Decentralized Object Location and Routing for Large-Scale Peer-to-Peer Systems, *18th IFIP/ACM International Conference on Distributed Systems Platforms (Middleware 2001)*, Heidelberg, Germany.
- Stoica I, Morris R, Karger D, Kaashoek M F and Balakrishnan H, 2001, Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications, *ACM SIGCOMM 2001*, San Diego, California, USA, 149–160.
- Botts M, Percivall G, Reed C, and Davidson J, 2008, OGC Sensor Web Enablement: Overview and High Level Architecture, *Lecture Notes In Computer Science*, vol. 4540, 175–190.
- Na A and Priest M, 2005, OpenGIS Sensor Observation Service Implementation Specification, *OGC Web Service Standard Specifications*
- Goodchild M and Yang S, 1992, A Hierarchical Data Structure for Global Geographic Information Systems. *Journal of Graphical Models and Image Processing (CVGIP)*, 54(1):31–44
- Deshpande A, Nath S, Gibbons P and Seshan S, 2003, Cache-and-query for wide area sensor databases. *ACM SIGMOD*.
- Ahmad Y and Nath S, 2008, COLR-Tree: Communication-Efficient Spatio-Temporal Indexing for a Sensor Data Web Portal, *IEEE Conference on Data Engineering (ICDE)*.