The Future of Mobile Survey Data Collection in Natural Resource Surveys

Sarah Nusser(*), Keith Clarke(+), Michael Goodchild(+), Leslie Miller(**)

* Department of Statistics
& Statistical Laboratory
** Department of Computer Science
Iowa State University
Ames, IA 50011-1210
tel: 515-294-9773 and 515-294-4377
fax: 515-294-2456
email: {nusser,miller}@iastate.edu

+ National Center for Geographic Information and Analysis and Department of Geography
University of California in Santa Barbara
Santa Barbara, CA 93106-4060
tel: 805-893-7961 and 805-893-8049
fax: 805-893-3146
email: kclarke@geog.ucsb.edu,
good@ncgia.ucsb.edu

www.statlab.iastate.edu/dg

Project Background

We present 4 posters that present the full range of our interdisciplinary Digital Government research program. The first poster describes a conceptual framework for accessing, using and collecting geospatial information in mobile data collection environments. The next 3 posters discuss specific components of this research, including interoperable digital geospatial libraries, wearable computing technologies for field data collection, and middleware to support adaptive exchange and analysis of geospatial data between the field computing and repository environments. In the final poster, testbed environments will be used to illustrate research principles in federal statistics applications.

Project Overview

The National Resources Inventory (NRI) is a federal statistical survey sponsored by the US Department of Agriculture to provide information on land-use and natural resource conditions and trends for nonfederal lands. Every year, survey data are collected through photo-interpretation of low-altitude photography from approximately 200,000 points distributed throughout the US. In any given year, approximately 40,000 of the photo-interpretation points will be visited to obtain additional field observations for subject matter variables and to quantify measurement error properties.

The NRI survey program has used handheld computers in the field since 1996 as the client in a client with dial-up access to central data server. A computer-assisted data collection form is used to collect and check data, and when in the field, an enhanced interface has been developed to support acquisition of GPS readings by naïve users. However, no graphical forms of geospatial data have been incorporated in this system due to limitations associated with storage and

bandwidth. It is of interest to incorporate access to digital imagery and other forms geospatial information resources within a field data collection system in a manner consistent with the framework proposed under Project Battuta.

A small NRI testbed is being established to demonstrate and elaborate on research being conducted to explore the proposed framework. The region for the testbed represents a 9 sq mi area where farmland and riparian areas are being developed for urban purposes. Mock sample units have been created for field data collection, along with historical data that simulate a typical NRI database for each sample unit. Repositories are being constructed and existing digital libraries have been identified that contain geospatial resources that would be useful for field activities. These include digital orthophotography, digital raster graphs, soils maps, and a sequence of aerial photographs of varying formats that show how land cover/use has changed since the mid 20th century. In the first phase of implementation, ad hoc spatial resources, such as road maps and other types of thematic information, are being identified to play the role of less structured geospatial data that is available via the web. Where possible, agent infrastructures are being constructed to demonstrate middleware functionalities. Future implementations with focus on fully adaptive mediation of ad hoc and prepared information resources. Field activities under consideration include campaign and route planning for multiple sample units, navigation to a specific sample unit, collecting GPS-determined coordinates and boundaries, using additional geospatial and other information resources to assist with interpreting conditions for the sample unit, and annotating orthophotographs to record conditions and geographic information at the site. Currently, the field device for the testbed is a palmtop computer with a wireless phone connection. As research progresses, a wearable computer with an integrated positioning system and continuous wireless access to digital geospatial information repositories will be incorporated into the testbed.

We will use the testbed to illustrate how Project Battuta research initiatives transform and expand field activities for natural resource and land-based surveys. Extensions of the NRI testbed to include other USDA and USGS settings will be discussed, as well as early work on establishing a testbed for demographic and economic surveys.

A New Framework for Computer-Assisted Data Collection

Sarah Nusser
Department of Statistics & Statistical Laboratory
Iowa State University

Research in mobile computing for survey data collection has been grown exponentially over the last five years, with the greatest emphasis placed on handheld and tablet devices for large-scale statistical surveys. Models for computer-assisted survey information collection are still largely based on the client-server paradigm of telephone survey interviewing labs. Little attention has been given to mediated information exchange, digital geospatial information resources, or emerging mobile environments such as wearable computers.

The goal of Project Battuta is to conduct research that harnesses the potential of evolving technologies and alternative information resources within the field data collection process. To

do so requires a new paradigm for field data collection that draws upon integrated and distributed geospatial information resources, modern computing paradigms for mediated infrastructures, and emerging mobile technologies into the computer-assisted field data collection environment.

Under our proposed framework, we assume that a field user requires access to digital repositories prepared by the agency sponsoring or conducting data collection, as well as less structured access to geospatial information resources more generally available via the Web. At the same time, we recognize that the field user may be technologically naïve and that the field interface may be extremely limited relatively to the large volumes of information that exist in distributed geospatial data repositories. Thus, the infrastructure must seamlessly negotiate for the user, hiding the complexities of interacting with prepared and ad hoc information resources and delivering data and information in forms appropriate for the field computing and application environment. A mediating infrastructure that understands the user environment and that can intelligently search for and prepare information requested by the user is essential to providing this seamless and context-aware interface.

The underlying model for the framework is one in which the field user is capable of accessing prepared stores as well as distributed information resources available via the Web. Middleware is used to intelligently interact with the user and data sources to effectively address user-specified needs. This includes not only query processing, but also the potential to generate local caches to temporarily store results and computation servers that perform intensive tasks such as conflation to prepare appropriate information products for the field user. Several parameters drive realizations of this model that are appropriate for a particular data collection setting. These include the user's physical and computing environment, network constraints, repository characteristics, and the types of tools and products required to support data collection activities in the field. The benefits and functionality created by this framework for will be illustrated with examples from statistical surveys and less structured applications such as criminal justice and crisis management.

A Prototype Wearable System for Field Computing

Keith Clarke
National Center for Geographic Information and Analysis and Department of Geography
University of California

As part of Project Battuta, the University of California Santa Barbara under the National Center for Geographic Information and Analysis has been researching how state-of-the-art wearable computing technology can be of use in field data collection. We are currently building a prototype of such system, using the CharmKit PC-104 system. The prototype includes input from a single-hand keyboard called a Twiddler; from a digital compass with roll, pitch and yaw tracking; and from a small GPS receiver. Output is delivered to a MicroOptical eyepiece, which clips onto a regular pair of glasses. This system will be described and demonstrated, and experience gained in its construction shared.

Further work with the prototype relates to modes of operation. Two modes will be implemented: a navigation mode that leads the user through geographic space, using prior paths, or computed

navigation information; and a positioning mode, where assistance to the user in geographic positioning and spatial alignment are given. Data collection will be conducted by filling in on-line Internet-based data base forms, presented to the user in context by Java applets running on a Mobile Internet platform. We will discuss research being conducted into aspects of the user interface such a system would require, including icons and graphics for the two modes, and map-based displays.

Using Geolibraries in the Field

Michael Goodchild
National Center for Geographic Information and Analysis and Department of Geography
University of California

In Project Battuta we are exploring the ability of personnel in the field to access massive archives of geospatial information. Such archives, known alternatively as clearinghouses, warehouses, and geolibraries, are growing in importance as Web-accessible stores of digital maps and images. The paper begins with several use cases, outlining situations in which field workers need to download such information in support of field activities, including finding and navigating to observation sites; registering maps and images to local GPS measurements; examining maps representing past conditions; and using maps and images to direct spatial sampling designs that are continuously modified in the field.

Much effort in the geospatial community is currently being expended on protocols in support of remote access to distributed geolibraries. All of these efforts aim to increase the functionality of clients, and to reduce the effort required to search, access, and extract geospatial data from archives. The paper reviews these efforts and evaluates them against the requirements of Battuta.

The paper ends with a demonstration of some specific use cases, including extraction of images from digital libraries; their registration in the field; and a simulation of adaptive spatial sampling in which data already collected are combined with ancillary data obtained from a digital library to redirect sampling.

An Infrastructure for Delivering Geospatial Data from Heterogeneous Data Sources to the Field

Leslie Miller
Department of Computer Science
Iowa State University

A key feature of any environment designed to give field workers access to geospatial data is the infrastructure used to connect the field devices to the data sources. The infrastructure must be very flexible in its ability to obtain data. At the same time, it must be capable of minimizing the amount of data flowing through the network. To do this, we have chosen to make use of object-oriented views implemented as mobile agents. The views provide an excellent basis for deriving the data for the user's request and the mobile agent aspect creates a great deal of flexibility in the location for integrating or analyzing data.

The implementation of our view agent infrastructure model makes use of wrappers, mediation, and XML. The wrappers are used for encapsulating the data sources and the mobile field devices. As is generally the case, the wrappers allow the details associated with the heterogeneity of the data source (or device) to be localized. The result is that within the boundaries of the wrappers, the mobile view agents work in a relatively homogeneous environment of manipulating XML encoded data.

The internal infrastructure environment is also populated with a set of computation servers. Each computation server has a local object-oriented data warehouse equipped with a set of tools designed to work with geospatial data. Since the prospect of query reuse is likely for a field worker, we store the final and intermediate results in the data warehouse, allowing the warehouse to act as an active cache.

The combination of these tools gives us a dynamic, adaptable infrastructure for handling geospatial data in field applications.