6-27-2005

Standardizing Geospatial Information for New England Conservation Lands: Perpetrual Data Maintenance and Distrubuted Data

Richard Sutton
Applied Geographics, Inc.

October 2005

Standardizing Geospatial Information for New England Conservation Lands

Perpetual Data Maintenance | Distributed Data Capture

The New England Environmental Finance Center
Edmund S. Muskie School of Public Service
University of Southern Maine

Richard Sutton
Applied Geographics, Inc.

CONTENTS

The New England Environmental Finance Center wishes to acknowledge and thank the numerous individuals who have assisted in this project. These include federal, state, local and land trust officials, private sector software and data vendors, planners and private citizens who have contributed their time and expertise as the project has evolved since preparation of the initial Conservation Lands Feasibility Study in 2002.

Ongoing project status as well as extensive analysis information may be accessed from:
http://www.appgeo.com/clients/efc/

Additional project information as well as other EFC initiatives may be accessed from:
http://efc.muskie.usm.maine.edu/

Executive Summary

An explosion is occurring in the access to and usability of geospatial data. This is being facilitated not only by traditional GIS giants such as Environmental Systems Research Institute (ESRI), but more dramatically through new web services offered by the likes of Microsoft, Google and Yahoo - the largest players in the software and Web landscape.

While these new services tend to be carefully and cleverly engineered with capabilities continually improving from software and content delivery enhancements, the actual quality of data being served through them is not keeping pace across many critical layers. High quality, responsive services are becoming available at low cost or even free of charge, yet few systems are currently in place to increase the quality of the content through input from the growing body of spatial data users. That is, few services and providers are even attempting to mine the knowledge of the wide array of participating individuals for their local expertise to increase the quality of these data resources over time.

A familiar pair of truisms of traditional, paper based cartography states that *no map ever goes far enough at its edges and none is ever as accurate as it should be in the local area that the user knows best*. The first of these weaknesses is being eliminated by fast web mapping services that cover the entire United States and in some cases the globe (and recently, Moon and Mars as well).

The second issue, inadequate local accuracy, will only be finally corrected by drawing on the knowledge of a large group of local experts and implementing a scheme of distributed data capture through which they may contribute efficiently to the repository. Harnessing this resource has not been addressed with anything approaching the same technical focus or imagination as content delivery.

This brief paper addresses the problem in microcosm as it occurs in the case of conservation lands data for the northeastern United States. Through an ongoing initiative with Applied Geographics in Boston, the New England Environmental Finance Center (NE/EFC) has worked to identify friction points and opportunities for increased efficiency in the conservation lands data capture and standardization process over the EPA Region 1 (New England) area. Like other thematic layers, conservation lands data are typically best captured as polygons which carry tabular attribution of varying complexity depending upon which state or organization collects and maintains them. By example, Massachusetts has collected information on more than 30,000 parcels and informed these polygons with a fully relational database that contains dozens of tables with nearly 100 active attribute fields. Maine is at the other extreme, and with four times the overall land area has barely one twentieth the number of cataloged conservation properties and a very restricted set of tabular data associated with them. Most of the properties that have fallen through the cracks in Maine belong to the municipal or land trust categories. These are prime candidates for distributed data capture, being broken into small jurisdictions where a large number of local experts have very clear knowledge of their own area but no easy means of passing this knowledge on to others working in a more regional, state or federal capacity.

This issue is widespread: the field of users of digitally stored and delivered geospatial data is growing at an unprecedented rate but there is little systematic effort being taken to leverage the local expertise of these users back into a unified and consistent data store. It is highly desirable that such a data store should grow so as to perpetually increase the accuracy and detail of the overall content. Unfortunately the data acquisition of many layers, including conservation lands, continues to depend upon large scale, centralized and expensive blanket initiatives that attempt to update all

features with one push, but suffer by being out of date almost as soon as they are delivered. The unfolding dynamics of the Web indicated that this will not be the case in the future. This paper suggests some steps that may be taken to accelerate the process.

Context

Protected open space defines a society, both by the strict boundaries it imposes on future development as well as through the nature, number and size of the properties that are secured. Land has traditionally been preserved in order to protect vital water resources, secure tracts for recreation and wildlife habitat, safeguard agricultural resources and spare aesthetic assets. Open space protection is necessarily a long term, evolving process.

Tracking these properties for purposes of inventory and analysis is obviously a long term process as well. Globally, nationally, regionally, even locally, different jurisdictions tasked with maintaining protected open space records approach the problem in varying ways, update them at different frequencies and distribute the content according to differing hierarchies of access and restriction.

Before the existence of geographic information systems there was no practical way to accurately evaluate large numbers of complex properties over wide, multi-jurisdictional areas. In fact there was little perceived need to do so. Land acquisitions were made locally and local considerations were in most cases all that mattered. But with the advent of cheap and ubiquitous computing power and data storage, and with more than three decades of investment sunk into spatial data technology by public and private innovators, the picture is very different in 2005.

Statewide layers describing wetlands, watersheds, aquifers, critical plant and animal habitats, as well as accurate base data including highly accurate transportation features and aerial imagery are now the norm rather than the exception. Many of the base layers have been consolidated to national consistency through public initiatives such as The National Map or are easily accessible through commercial channels provided by Microsoft, Google and Yahoo.

This rich access to accurate spatial data fuels greater needs and expectations for quality and availability over a wide user base. As the market penetration of such (currently free) products as Google Earth and VirtualEarth from MSN makes roads and aerial imagery globally available at sub-meter accuracies, users come to expect vector-based thematic data of comparable quality. Standard planning exercises such as buildout analyses and suitability siting studies are conducted in GIS environments using these base layers as starting points. When thematic vector layers such as conservation lands do not match these accuracy levels even the most novice users will easily detect the discrepancies.

But for many of these thematic layers, including conservation lands, there has been little if any effort to standardize capture and distribution practices among groups of states or nationally. In addition, the accuracy of the data is highly variable between the different entities who maintain and distribute them.

Much of the standardization burden surrounding data types and spatial projections has been alleviated by software advances in recent years. ESRI deployment of on-the-fly projections at ArcView 8.x and later releases, as well as their versatile interoperability functionality co-opted from Safe Software has made some of the most tedious and ‘black art’ aspects of cross boundary/cross format data integration transparent and effortless.

But the software can only go so far toward ensuring quality and interoperability. Ultimately the quality of the data being delivered through slick, fully Web-enabled services is only as good as the individual spatial features and attributes from the tens of thousands of locations comprising them.

Constraints and Opportunities

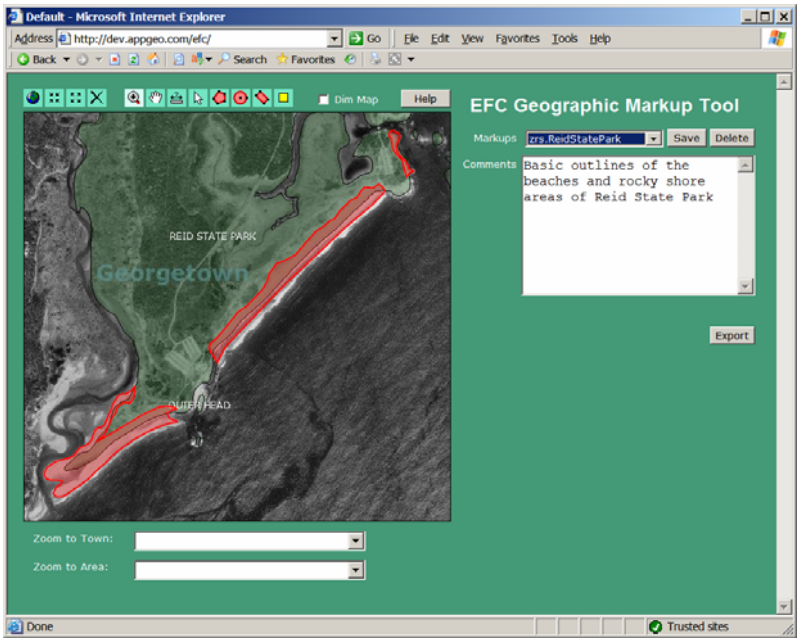
Collaborative Data Capture

Extensive discussions with data providers and users have identified two primary pinch points in the overall process. The first of these involves accessing and organizing the large number of stakeholders and local experts that need to participate in ongoing data maintenance. The needs and technical capabilities of these many individuals vary widely, and their data requirements are by no means uniform. Land trusts are increasingly interested in seeing an accurate picture of property holdings in their areas of operation, for purposes of management planning and acquisition strategy. Municipalities use protected lands data to assess future build patterns and growth limitations. State agencies depend on these data for accurately integrating with large area wildlife and recreation planning, and the federal government and multi-state agencies require inter-jurisdictional consistency to evaluate future protection or asset liquidation involving vast watersheds or bioregions.

No means currently exists to draw all of the expertise of these disparate stakeholders into a single, cohesive data set covering New England. In fact, many areas suffer data inaccuracy all the way down to the municipal level, such that it is not possible to interrogate any existing data set to produce an accurate picture of protected holdings even within a single town.

One of the approaches THE EFC has investigated for alleviating some of these data holes is the production of a Web-based geographic markup tool.

The strategy of deploying such a tool is to make feature capture of local specifics free and simple to as large an audience as possible across the entire New England landscape.



The *EFC Geographic Markup Tool* graphic displayed at right shows a sample screen of this application. Using a standard Internet browser with a small plug-in, users can display base imagery and scribe feature shapes over them.

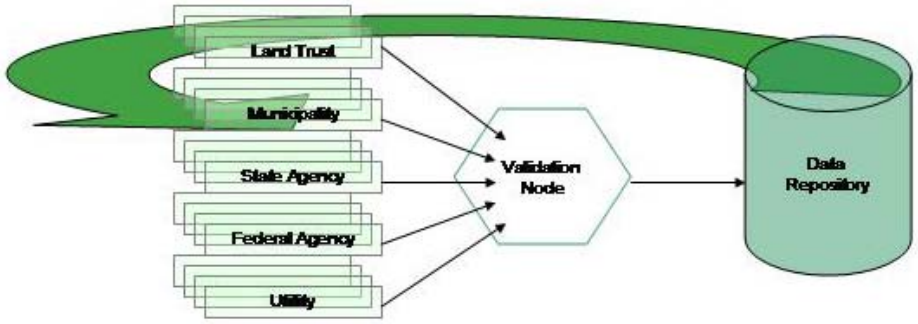
Base web services are provided by state repositories, in this case Maine Office of GIS. Conservation features are consolidated from respective states into a composite that is hosted by

EFC/AppGeo, and are fused with supporting services by the AppGeo ArcIMS web service. The user sees only a single map, despite the complexity and numerous sources of the contributing data.

Users of the tool are able to save their marked sketches as named entities within the application as well as export them to ESRI Shapefile format for exchange and integration in other GIS packages. This provides as simple and accessible a solution as possible to provide a spatial data capture option for local experts without access to editable GIS technology. However, it is important to note that this is a demonstration application deployed as proof of concept rather than a field hardened automation tool. The application will require further enhancements and user interface modifications to perform reliably in a high volume production environment.

Data Validation

The second bottleneck occurs later in the data production chain, at the point where the data are inspected, qualified and validated into a repository for storage and distribution. Ironically, this pinch point is and will be increasingly exacerbated by the success at creating an efficient capture apparatus up front.



Evidence of the success at distributed data capture is apparent in Massachusetts at MassGIS where the backlog of data pouring in for inspection and check-in through the single validation point has produced a considerable backlog.

Of course in Massachusetts the data is already accurate down to very small (sub-acre) parcels in many areas throughout the state. Massachusetts has taken serious steps to leverage the newest ESRI technology into a relational database structure that efficiently normalizes the large number of attributes and to streamline the validation process as much as the technology will permit. Yet their update process is still hostage to many paper records changing hands and a large number of file cabinets constituting their archives.

Variable Data Access and Content Confidentiality

Different stakeholders require different data access and restrictions to the content over which they have stewardship. Different property owners, managers and easement administrators express very different ideas about what they feel is appropriate privacy regarding information about their lands. Land trusts tend to be the most conservative in this regard, more so in the northern states than the southern tier. Most protected land records, whether fee owned or easement properties, are fully

disclosed through publicly available GIS data in Massachusetts. Rhode Island and Connecticut tend to follow this pattern though there data inventories are not in nearly as complete shape as those of Massachusetts. Land trusts in the northern states are much more restrictive about public release of data pertaining to these properties. There proprietary approach to conserved lands data is frequently based on the notion that making the data public will invite a deluge of unwelcome trespassers.

In the southern New England states where land is far more urbanized and protected open space properties are much smaller (but massively more numerous) there is significantly less sense that unimpeded access to the data will invite unwelcome guests. It seems that trespass stress – the excessive utilization of land by uninvited and unwelcome users – is a potent indicator of an inadequate quantity of publicly accessible land in a given area. An accurate accounting of protected lands analyzed together with demographics should provide an indication of where trespass stresses will develop. The remedy for alleviating such stresses in these areas should be a campaign to secure more land, not one to close down access to the data.

Nevertheless, it will never be appropriate to try to open up all data to everyone all of the time. Any unified conservation lands data system must allow individual stakeholders to protect data that they deem too sensitive to make generally available. Data describing properties such as non-public conservation easements designed to protect relatively small parcels and provide tax relief are important to the work of local planners and trust officials, but probably do not need to be made widely accessible. Information about such properties is public in the sense that it can be uncovered in recorded documents in deed registries. Easy GIS access potentially increases speculation vulnerability.

Developing an access hierarchy that respects the privacy of specific contributors while passing as much of their content as possible into the Region 1 composite repository is a tricky exercise and no clear set of rules exists to define where the lines should be drawn. That the data must support analysis from the bottom up as well as the top down is an ambitious objective; providing sufficient incentive to land trusts to convince them to share their data with a potentially large number of anonymous data consumers is another matter entirely. But there is no question that technical tools and GIS assistance directed into many of these organizations will help them to further conservation goals even if significant portions of their data don't make it out of the land trust community.

Costs and Funding Candidates

Scale and Perspective

Annual costs of maintaining conservation lands data across EPA Region 1 are difficult to compile, as records are not updated in an ongoing, continuous fashion; many participants are involved in varied and subtle capacities; and there is wide variation in practices across the six state area. There is also a considerable amount of subsidized and volunteer work undertaken by concerned citizens, land trust members and after-hours conservation commission officials that must be considered if costs are to be fully and accurately accounted.

The dynamism and complexity of these records dictates approximately a full time equivalent staff person per state just to properly manage the validation node. As described above, in some states this is level of effort is already proving to be largely inadequate.

Organizations such as The Nature Conservancy and larger local trusts spend significant staff resources compiling spatial data pertaining to their own and collateral assets, and many of these make their way into public repositories at the state level. With so many contributors and stakeholders involved in the effort, an estimate of $1 million per year as a base cost for keeping conservation lands records consistent and current over the Region 1 area likely understates actual costs.

If this seems high, consider that all told more than 50,000 parcels of land in New England are under some form of protection – government ownership, trust controlled fee ownership or some easement mechanism - against future development. The intangible value of these lands is incalculable, but using a "highest, best use" metric it would total billions of dollars. Assessed in this light, $20/parcel it does not seem such a high price to annually review data status to maintain a proper accounting of these resources.

Potential Funding Initiatives

The cost of continuing to push the different conservation lands data streams into closer alignment will be comparatively very low. Small investments in appropriate areas may be used to tighten integration of capture and distribution processes among capture entities and state level validation officials. Short descriptions of some of these follow.

- Provide seed funding to states to ensure that all Region 1 web services work together. This should be an easy modification for states with existing web services. Costs could probably be held under $5,000/state
- Provide financial support or direct intern assignments to ease pressure on validation node bottlenecks at state level data centers.
- Work with land trusts to pool ESRI and other spatial software buys/licensing in order to minimize product outlays and reduce aggregate purchasing/leasing
- Continue to support EPA Region 1 conservation lands portal to simplify data access for interstate/wide area data consumers
- Support ongoing property parcel automation in the interest of harvesting protected open space features from these data sets
- Encourage buildout analysis efforts and development of municipal/watershed scale analytics that create demand pull for more accurate conservation lands data.
- Foster ongoing forum discussions among core data stewards, technicians and large consumers of these data throughout the region to synchronize efforts, exchange best practices and foster interpersonal trust among participants.

About the NE/EFC:
The purpose of the New England Environmental Finance Center is to further the joint goals of the U.S. EPA and the Muskie School of researching, publishing, and extending creative approaches to environmental protection and management, especially respecting the associated "how-to-pay" questions. In particular, the Center works to advance the understanding and practice of "smart growth" throughout New England; to build local capacity to deal with related issues; and to develop and apply techniques that go "beyond compliance" with government regulations.

For additional information contact:
Samuel Merrill | Director, Assistant Research Professor
New England Environmental Finance Center | Edmund S. Muskie School of Public Service
University of Southern Maine
49 Exeter Street, #205 | Portland, Maine 04104
(207) 228-8596 | smerrill@usm.maine.edu

Richard Sutton
Applied Geographics, Inc.
Boston, Massachusetts
(617) 259-1690 | rs@appgeo.com

Edmund S. Muskie School of Public Service
University of Southern Maine

Bulk Rate
US Postage
PAID
Permit 1234
Portland, ME