Biol Invasions (2007) 9:597–599
DOI 10.1007/s10530-006-9042-4

INNOVATION NOTE

Balancing data sharing requirements for analyses with data sensitivity

Catherine S. Jarnevich · James J. Graham · Gregory J. Newman · Alycia W. Crall · Thomas J. Stohlgren

Received: 16 June 2006 / Accepted: 12 July 2006 / Published online: 14 November 2006

Abstract Data sensitivity can pose a formidable barrier to data sharing. Knowledge of species current distributions from data sharing is critical for the creation of watch lists and an early warning/rapid response system and for model generation for the spread of invasive species. We have created an on-line system to synthesize disparate datasets of non-native species locations that includes a mechanism to account for data sensitivity. Data contributors are able to mark their data as sensitive. This data is then 'fuzzed' in mapping applications and downloaded files to quarter-quadrangle grid cells, but the actual locations are available for analyses. We propose that this system overcomes the hurdles to data sharing posed by sensitive data.

Keywords Exotic species · Invasive species · Sensitive data · Data sharing · Data synergy

C. S. Jarnevich (✉) · T. J. Stohlgren
U.S. Geological Survey, Fort Collins Science Center, 2150 Centre Ave Building C, Fort Collins, CO 80526, USA
e-mail: catherine_jarnevich@usgs.gov

J. J. Graham · G. J. Newman · A. W. Crall
Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO 80523, USA

The invasion problem

Invasive species present one of the greatest challenges in the 21st century due to harmful effects to native species and ecosystems, human health, and the economy (Mack et al. 2000; Pimentel et al. 2005). Additionally, costs of control drastically increase and the possibility of eradication drastically decreases with the size of the infestation (Rejmanek and Pitcairn 2002). Thus, we need coordinated efforts for early detection and rapid response to contain new infestations of invasive species. Likewise, it is necessary for resource managers to have a mechanism to share information on locations of invasive species, which cross political and management boundaries. As simple as it sounds, several cultural and technological issues constrain the sharing of data.

Constraints to data sharing

There are at least three cases involving species data where the data could be regarded as ''sensitive'', and barriers to data sharing could exist despite the necessity of merging the sensitive data with other data for analyses. The United States is a patchwork of land ownership, mixing areas of private, city, county, state, and federal lands. Information for species found on public land is generally in the

public domain, but private property rights' issues arise when considering knowledge of species found on private lands. For example, some counties can cite private land owners for failure to control noxious weeds. The weeds themselves can often lead to a devaluation of the property, such as the identification of leafy spurge (*Euphorbia esula*), a species that can be toxic to cattle, on a ranch in western North Dakota.

Openly sharing data can lead to poaching. For example, the potential poaching of biological control agents from sites where they have been released is especially worrisome in their initial phase where the cost for organisms can be very great (e.g., $1/each). Poaching at pilot biological control sites can have devastating consequences if the pilot population is not yet established enough to withstand harvesting, whereas if the population were left alone for a certain amount of time, it would probably develop into a good source for collection.

A final, tangential area of sensitive data involves location information on rare, threatened, or endangered species. In analyzing the impact of invasive species, it can be helpful to examine their distribution in relation to sensitive species. However, locations of these sensitive species are restricted to agency personnel to help protect them. Poaching can have harmful effects on local populations of these rare species just as it can on establishing biological control sites.

What data are sensitive

Addressing problems of sensitive data requires first ranking their sensitivity. The most sensitive data could be related to issues of rarity, protected status, or ownership, as described above. However, we asked data providers if it was actually the species, the biological control agent, or the location that was sensitive. We determined that it is actually the location that is sensitive for both species and biological control agents. The exact locations of other organisms such as common non-game species or non-native species found on public lands are generally less sensitive than the locations of game species, commercially valuable species, non-native species on private lands, etc.

Our solution

We have been developing the website for the National Institute of Invasive Species Science (http://www.NIISS.org) that is built on the Global Organism Detection and Monitoring system (GODM). This organization is a U.S. Geological Survey led consortium of federal to local agencies and organizations that is intended as a forum for resource managers to share information on the location, abundance, and control measures for species. During this development, we were required to create a method for handling sensitive data that would still facilitate the sharing of information.

There are several steps to be considered in securing sensitive data. Important factors include: (1) having a secure login system; (2) having a way to categorize data as sensitive; and (3) effectively managing sensitive data. Therefore, we have taken several steps to implement sensitive data security. First, our website allows different levels of access to website pages and functionality based on a user's login. Guests (the open user community) can visit the website, but will have limited access. Registered users may request to be approved for higher levels of access. These higher levels include the ability to contribute and manage data, including sensitive data, to the website.

Second, before any registered user can contribute data to the website database, they are required to agree to a data sharing agreement. One of the eight points in the agreement states: ''The data provider warrants that they have made the necessary agreements with the original owners of the data and have the proper permissions so that they can make the data available through the National Institute of Invasive Species Science database and website.'' This statement is to ensure that the contributor of the data has permission for the data to be posted on the website from the original data collectors. For example, someone wanting to post a location in their neighbor's yard would need permission from their neighbor to post information about their neighbor's private property.

The third step taken to ensure sensitive data security occurs with the actual data upload process. All data contributed to the website are as-

signed to a ''project''. Projects are created and managed by website users, and the person who created the project controls what other users belong to the project and therefore can contribute data to the project (i.e., be a project contributor). During the data upload process, the contributor is asked if any of the data being added to the website database needs to be marked as sensitive. If so, they are given four choices for 'data fuzzing' including 7.5, 30 min, 1/2 °, or 1° grids. The 7.5 min gridding is currently implemented and the others will be added in the future. We have systematic grids for the United States at these different resolutions. The data point will be assigned to the grid cell it falls within, rather than generating a grid cell around the data point with it as the bull's eye. Restricting access to exact locations of sensitive species or biological control agents also solves other problems dealing with ways people could obtain the sensitive data. For example, the savvy website user is thus unable to determine the location of the sensitive species by obtaining what other species were found with the sensitive one and obtaining a location for those collocated non-sensitive organisms through a species search on the website, as the location of the collocated organisms would be sensitive.

Data fuzzing is implemented based on a website user's login and project membership. A person must be logged into the website and belong to the data's project to view and download exact sensitive data locations. Other website users will see the 'fuzzed' locations in the mapping application or in a data spreadsheet and can only download the 'fuzzed' locations (e.g., location as a quarter-quad polygon if 7.5 min quadrangles were chosen as the sensitivity level rather than a single point).

The above steps ensure that the sensitive data locations will be protected. However, recording the exact locations in the website database is very important for analyses. While the sensitive data locations are not provided to non-project members, the actual data are used for any users performing analyses such as simple statistics or a habitat suitability model for the species with sensitive data. If the fuzzed locations (say at 1° grid) were used in modeling, it would not be very useful because the area is too large to determine precise species–environment relationships with other data layers. By using the exact location in modeling, more information will be gained from the environmental variables associated with the sensitive locations. Thus, the exact locations can still provide valuable input to models where the results and not the original data are shown.

Conclusion

We propose that this solution balances the need to share knowledge of invasive species locations against the need to retain location sensitivity. We are able to create distribution models and watch lists for areas using sensitive data, while providing locations at resolutions (e.g., quarter-quad) that protect exact coordinates. These methods may prove useful for other types of sensitive information where specific locations are confidential.

Acknowledgements We would like to thank the U.S. Geological Survey Fort Collins Science Center and Colorado State University Natural Resource Ecology Laboratory for logistical support. We would also like to thank our early users of the website who made us aware of these issues. To all we are grateful.

References

Mack RN, Simberloff D, Lonsdale WM, et al (2000) Biotic invasions: causes, epidemiology, global consequences, and control. Ecol Appl 10:689–710

Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecol Econ 52:273–288

Rejmanek M, Pitcairn MJ (2002) When is eradication of exotic pest plants a realistic goal? In: Veitch CR, Clout MN (eds) Turning the tide: the eradication of invasive species. IUCN SSC Invasive Species Specialist Group, IUCN, Gland, Switzerland and Cambridge, UK