TITLE:
GIS Data Citation Rates: are data being properly credited in lists of references?

Author:
Kristen B. LaBonte
Life & Environmental Sciences Librarian
University of California, Santa Barbara
Santa Barbara, CA 93101-9010
klabonte@library.ucsb.edu

ABSTRACT:
Researchers in the field of Geographic Information Systems (GIS) use large data sets, many of which are combined with additional data sets to analyze and display information. Data relating to research is usually collected in the field and combined with layers of data from other sources. These sources are often credited within the text of an academic paper, but are not reported in the list of references. Citation of data is important in order to acknowledge and validate the source, and to create ease of access for other users. This study assesses the rates of GIS data citation in a sample of peer-reviewed academic journal articles from the years 2002 and 2012 to illuminate trends in citation patterns.

Body of Paper:
GIS maps are built with layers of information that often come from other sources. These layers of information are citable, yet many are frequently absent from the list of references or the text of the article. It is important that this information is cited for multiple reasons: these data may have errors; others may wish to use these data; and authors deserve credit for their work. It also assists in the peer review process allowing for information to be traced to its source and validated if needed.

Occasionally an article with a published map will have been created only with the authors' data. Typically though, downloadable background layers are used (such as roads, rivers, and digital elevation models) in addition to the authors' data. These layers are readily available from institutions such as the United States Geological Survey, universities, municipal governments, and ESRI, the makers of the most commonly used GIS program, ArcGIS. Many institutions, for example the Harvard University Graduate School of Design, have created finding aids for spatial data in their collections (Cote, n.d.).

A search of the *Web of Science, Aquatic Sciences and Fisheries Abstracts, Academic Search Complete, Biosis,* and *Google Scholar* on the following topics yields no relevant results to determine citation rates of GIS data, but does yield information relating to the importance of data citation and annotation:
("gis data" OR "geospatial data") AND (citation* OR citing OR cite).

For example, Mazzetti, Nativi, & Caron (2009) describe applications for geospatial data sharing for existing geospatial services and the need for the use of international interoperability standards for database backend structure for services, including data citation. As more GIS data is stored in repositories, there is a great need to incorporate curation principles into systems to ensure accurate citation over time (Bose & McGarva, 2007).

A further search on the same topics in *Google* yields numerous research guides and citations of data sets for use in lists of references. The first 10 results yield six academic web pages, a government page, and three from commerce. The first three pages of results contain relevant sites aimed at citing GIS and geospatial data. George Mason University Libraries have a library guide for Geospatial Data & GIS Services with a page specifically for GIS Writing & Citing, linking to other resources that are freely available online. Although information on how to properly cite GIS data is readily available, sources are still rarely cited in lists of references. Many databases, like *Web of Science*, now supply lists of references that are attached to articles. If GIS data is not cited there, data creators are not receiving proper credit for their work. This is an issue with many types of data sets. If proper citation occurs, data creators have the "potential to become an accountable part of the scholarly communication process" (Mayernik, 2012). A "Declaration of Data Citation Principles" is currently being drafted at The Future of Research Communications and e-Scholarship, FORCE11 site. Additionally, as data is accepted as a citable element of scholarly research, more papers will include data in the list of references.

In *Web of Science*, the search Topic = ((marine OR aquatic) and (GIS OR "geographic information system*")) yields 110 results from 2012 and the same search yields 41 results from 2002. These two sets of results were analyzed to determine if the citation trend is increasing or decreasing. The results were narrowed down to papers that actually use GIS data (determined by a printed map within the article) and further examined to determine if data was cited anywhere in the lists of references, inside the text of the paper, in the acknowledgments, or not at all. The materials and methods, acknowledgements, and reference sections of each paper were analyzed and specific keywords were searched. Those keywords were GIS, ArcMAP, "spatial data", "geospatial data", data, "Fig. 1" (or number corresponding with the map), bathymetry, and layer. The sample size started with 41 articles from each year, using a random number generator to for the 2012 articles, and was narrowed down to 26 articles from 2002 and 30 articles from 2012.

More than half of all articles examined had GIS data cited within the article as citations in the references, in the text, or in the acknowledgments. 58% of the articles from 2002 and 55% of the articles from 2012 had these citations. The citations were listed within the references in 23% of the 2002 articles and 20% of the 2012 articles. While these figures do not support an upward trend in citations, the sample size was not large.

This study should be replicated with a larger sample. For example, the same search could be conducted using *Aquatic Sciences and Fisheries Abstracts* instead of *Web of Science* because it is more specific to the field of marine GIS. Using the same search yields 117 peer reviewed articles from 2012 and 142 from 2002 in *Aquatic Sciences and Fisheries Abstracts*. Rather than using the same number of articles from both years, all articles should be examined and compared. The final sample sizes would be determined by the use of GIS after each article is examined. Additionally, articles that cite data within references can be examined to see if the publishers have an author requirement for the citation of data.

While a temporal trend in GIS data citations was not found, it is shown that almost half of the articles examined do not have data cited anywhere in the articles. This is problematic for the validation of authors' results, and the acknowledgments of others work. Over time, this problem should decrease as data is considered a valued element of research, as can be demonstrated through the rise of data management plan requirements of grant-funding agencies.

REFERENCES:

Bose, R. & McGarva, G. (2007). Safeguarding the Citation Lifecycle for Global Geospatial Repositories. *Geophysical Research Abstracts,* 9, 08548.

Cote, Paul. Cultivating Spatial Intelligence: Sources of Spatial Data. *Harvard University Graduate School of Design*. Retrieved November 27, 2013, from http://www.gsd.harvard.edu/gis/manual/data/index.htm

George Mason University Libraries. (2012). GIS Writing & Citing. George Mason University Libraries. Retrieved November 27, 2013, from http://infoguides.gmu.edu/content.php?pid=11647&sid=142993

Mayernik, Matthew S. (2012). Data Citation Initiatives and Issues. *ASIS&T Bulletin.* Retrieved November 27, 2013, from http://www.asis.org/Bulletin/Jun-12/JunJul12_MayernikDataCitation.html

Mazzetti, P., Nativi, S., & Caron, J. (2009). RESTful Implementation of Geospatial Services for Earth and Space Science Applications. *International Journal of Digital Earth, 2,* 40-61. doi:10.1080/17538940902866153

The Future of Research Communications and e-Scholarship. (2013) "DRAFT – Declaration of Data Citation Principles." Retrieved November 27, 2013, from http://www.force11.org/datacitation