



Supporting paper for the Expert Group on the Integration of Statistical and Geospatial Information, May 2015

Interoperability of statistical and geospatial data and metadata standards

Background

The Terms of Reference for the UN Expert Group on the Integration of Statistical and Geospatial Information (UN EG-ISGI) includes the following as a focus area:

“Encourage the use of existing, and development of new data and metadata standards as well as other standards to enhance the interface of location-based datasets from multiple sources.”

The Australian Bureau of Statistics has recently examined the issues associated with this focus area, and would like to share this work (see attached paper) for consideration by the Expert Group.

Introduction

The challenge of integrating, or joining, statistical data and geospatial data remains a significant barrier to the use of these data in analysis and visualisation, supporting business processes, and creating further value-added data. Barriers to data integration prevent the complete use of these data. Where integration is attempted, common experience is that 80%-90% of time is spent preparing the data and only 10%-20% of the time is used undertaking analysis, which wastes valuable resources. This wasteful preparation is also repeated across and within organisations. Moreover, manual manipulation by users who are unfamiliar with statistical or geographic data standards is highly prone to the introduction of errors that undermine the value of the subsequent analysis.

Current Challenges

As with many problems, there are varying layers in the challenges that affect the integrating or joining of statistical and geospatial data.

At the simplest level, there is little consistency in the way in which statistical metadata is structured to link to specific geometry (the primary geospatial metadata). Principally this statistical metadata are the codes and labels that notionally act as the link to boundaries or locations represented by geometries in geospatial datasets. Currently there is no physical, machine actionable, link between the code in the statistical dataset and the geometry associated with that code. Structuring this metadata more consistently will improve the ability to join data, both through direct human intervention and in more automated processes, such as machine to machine interactions.

At the next level up, there is a need to substantially improve the interoperability of statistical and geospatial data and metadata. This will allow for machine to machine interactions and the seamless integration of data within applications. This interoperability can be enabled through the availability and use of agreed data and metadata structures (such as the ISO 19100 series of geospatial standards and the statistically orientated Data Documentation Initiative – DDI), and through the use of common implementation standards (such as the OGC Web Map Service Interface Standard – WMS or the Statistical Data and Metadata eXchange – SDMX – web service guidelines). The use of such common elements will facilitate the communication of information about the data and enables its exchange and integration across platforms and data sources.

At yet another level, there is a need to enhance the understanding of the commonalities and differences in the information models used by the geospatial and statistical communities. This would inform the future development of data and metadata standards and the common implementation such standards. One particular issue that should be addressed is the different data models used in the two communities, where the statistical community uses multidimensional data models and the geospatial community generally use flat data structures. (This is discussed further in the attached paper.)

Mapping a way forward

To provide a mechanism for the Expert Group to progress the issues outlined above and to pursue the focus area of developing new data and metadata standards, the following recommendations are proposed:

- 1) A small working group be formed to work on the data and metadata interoperability issues discussed in this paper, and related issues.
 - a. The working group should comprise representatives of Expert Group member states who are willing to actively contribute to the work required to progress these issues.
 - b. The working group should have representatives from both a statistical and geospatial perspective.
 - c. Other interested parties from the relevant standards organisations should be notified of the formation of the working group and may join if they are willing to actively contribute.
 - d. The working group should develop a terms of reference and work program for consideration and endorsement by the Expert Group (UN EG-ISGI).
 - e. The working group operate on a virtual basis rather than requiring face-to-face meetings.
 - f. The working group report its findings and recommendations progressively to the Expert Group (UN EG-ISGI), and other relevant standards groups with a stake in the activities of the working group.

The expert group could discuss these recommendations and endorse as appropriate.