A Statistical Spatial Framework to Inform Regional Statistics

Martin Brady & Gemma Van Halderen
Australian Bureau of Statistics, Canberra, Australia
Corresponding Author: m.brady@abs.gov.au

Abstract

Statisticians are increasingly being asked to link their statistical information to a location. This demand is coming from most stakeholder communities at local, regional, national and international levels. Many statistical organisations are responding to this demand, but are they providing this geospatially enabled information in a way that maximises the benefit for all their stakeholders?

The issue is not one of making statistics available within a geospatial context, but one of providing a consistent, repeatable and extensible approach to meeting this growing demand for geospatially linked statistics.

This paper will examine capability that the Australian Bureau of Statistics has developed to provide a consistent approach to spatially enabling statistical information to better inform the stakeholder community. The paper will describe how the application of the ABS developed Statistical Spatial Framework (SSF) will assist any organisation integrate their own and other statistical information that conforms to the SSF. Simplifying the integration of statistics from multiple organisations will provide richer, more comprehensive information to better inform decision makers.

Key Words: geospatial, statistics, geography, big data, spatial data

Statistical Spatial Framework

Linking people and economic information to a location

1. Introduction

> "The work on global geospatial information management over the past two to three years has confirmed that one of the key challenges is a better integration of geospatial and statistical information as a basis for sound and evidence-based decision-making."
> - Secretary-General, UN Economic and Social Council (2012)

The demand for linking statistical information to a location is increasing rapidly. In a recent UN Statistics Commission Programme Review, led by the Australian Bureau of Statistics, most National Statistical Offices identified that there was significant growth in the demand for geospatially (location) enabled statistics. Further, this demand was coming from all sectors: government, commercial, research and academic communities, and the general public.

This growth in demand for geospatially enabled statistics has occurred because people and organisations in these sectors see the benefits that arise from linking information to a specific location. Fundamentally they see the benefits in being able to answer questions about local communities such as:

- Who lives in this area?
- What businesses or services are here?
- What is happening in this particular location?
- How do the people and the things that happen in this area compare with other areas? How are they similar or different?
- What has changed over time in this area?
- What might this area look like in the future?
- Does this region require a similar or different level of services to another region? Is this likely to change?

Providing geospatially enabled information from statistical and administrative data sources can help to answer these questions. This information will allow people and organisations to better understand what is occurring at the local level and make more informed decisions at both the local and macro levels. Governments need to find out about pockets of disadvantaged, improve the focus of their service delivery and provide services to those that really need them, as well as making broader level policy decisions. Organisations in the commercial sector wish to expand their markets, sell more goods or services, adapt to change, and reduce costs. Local communities and individuals seek to engage with these organisations to ensure that their communities, cities, regions and nations are good places to live.

Internationally, organisations such as the UN, national governments and the NGO sector, are seeking location focused information that will assist them in better targeting assistance within countries. This information is increasingly being applied to improve support for the response and recovery efforts for natural disasters, and many other activities that require the use of their increasingly scarce resources.

2. Drivers increasing the demand for location content

The demand for location based information has been expanding rapidly at local, national and international levels and is being encouraged by a range of drivers. These drivers include:

- new technologies, such as Google and other online mapping, that are increasing consumer expectations and knowledge
- consumer based technologies that support location capabilities, such as: mobile phones, tablets and other mobile technologies that incorporate GPS technologies; Cloud data storage; Software as a Service
- business intelligence and social policy research that incorporates locational elements as key inputs and outputs is increasingly being relied upon as vital information for business and government decision making
- research activities are applying location to an increasing number of projects to improve the quality and relevance of research outcomes.

3. Location based business intelligence

Governments and business, through their daily business activities capture huge amounts of information that has, or could have, an associated location element. In addition, social networking is providing another, less structured, source of people related information (such as needs, desires, complaints and requirements) that is often location enabled. This large volume of information is central to the 'Big Data' issue and underpins the increasing importance of business intelligence in the day-to-day operations of government and commercial organisations.

In order to further enhance the client and transactional data they hold, many organisations are seeking effective ways to apply a location element to this data. They see that adding location context to their information provides significant value to their business intelligence and helps to reduce cost, reduce fraud, provide more focussed (and therefore profitable) products, improve their effectiveness and efficiency in delivering services, and for government it creates more relevant information to support the development of evidence based policies.

However, this 'new' location based business intelligence remains elusive for many organisations. Many find it difficult to access the capabilities required to effectively transform address and other locational information into useful geospatial information. The absence of clear frameworks in this area also makes it more difficult to successfully implement solutions. Many of these organisations are also not traditional statistical organisations, such as the 'people' focussed government agencies (e.g. Health, Social Welfare and Education) and many of the commercial organisations. These organisations may not be able to confidently address the legislative privacy and confidentiality constraints on the use of their information holdings. Moreover, they may have concerns with navigating the privacy expectations of the clients who provide the underlying information.

4. Location enhanced public data sets

The emergence of the open data movement offers great promise for increasing the amount of information available for public consumption. However, geospatially enabled datasets face a number of barriers to becoming publicly accessible sources of information. For these datasets, privacy and confidentiality are a particular concern as the location information potentially provides a very direct way of identifying individual people, households or businesses within these datasets. Government and business organisations are generally very aware of these issues and this has created a barrier to releasing these datasets outside the specific organisations that collect the information.

A fundamental challenge in creating publicly available datasets is to provide clear frameworks and methods for effectively managing the legitimate privacy and confidentiality requirements, while also unlocking the benefits arising from linking socio-economic information to a location.
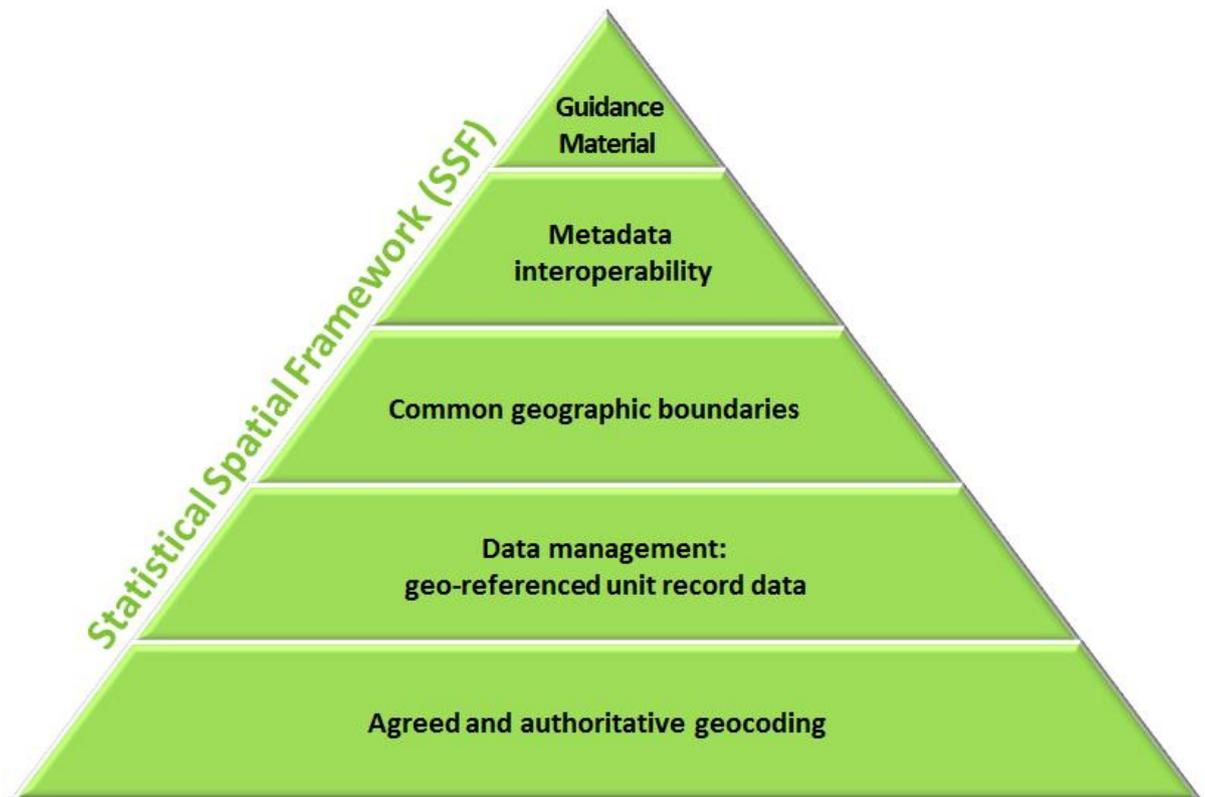
5. Statistical Spatial Framework

The Australian Bureau of Statistics (ABS) has recognised the challenges associated with applying a locational context to socio-economic information, as outlined earlier in this paper. Building on its significant history of geospatially enabling socio-economic information, the ABS has begun to develop the **Statistical Spatial Framework**. The ABS believes this Framework can provide the structured link between socio-economic information and location.

The Statistical Spatial Framework establishes a methodology to provide consistency when linking socio-economic information to a specific location, while also providing useful guidance material that addresses the complex issues of privacy and confidentiality and other related topics.

6. What is the Statistical Spatial Framework?

The Statistical Spatial Framework is a relatively simple concept consisting of five main components that are considered essential to integrating geospatial and socio-economic information. These are shown in the diagram below.

Statistical Spatial Framework Components

6.1 Geocoding

The Framework requires a common and consistent approach to establish a location (or geo-code) for each unit in a dataset, such as a person, household or business. This location should ideally be based on a physical address. Ideally, the geocoding should provide latitude and longitude for each unit level record in a dataset, since this will provide the greatest flexibility for future analysis and presentation. If this is neither practical nor cost effective, then another formal location reference system could be used. However, the geography used in an alternative reference system should ideally be persistent to meet future uses. In Australia, the National Address Management Framework (NAMF) will provide this commonality in address geocoding.

6.2 Data Management

The Framework recommends that the linking of a geo-reference for each person, household or business record in a dataset occur within a data management environment. For unit record level data, this means that the location geocode established during geocoding should ideally be stored so that it can be directly associated with each unit record. In addition, a geo-referencing geography, such as a Mesh Block or other small area geography, should be similarly stored for each unit record. Storage of location data at this level should occur regardless of whether the data will ever be publically released at the unit level. A high precision geo-reference enables any geographic context to be applied when preparing the data for release in the future (i.e. in aggregating data into larger geographic units).

This component of the Framework also recommends that established data management, tools techniques and standards facilitate the integration and management of the geo-reference within the dataset.

6.3 Common Geography

To provide the ability to make comparisons across datasets from different sources, the Framework recommends that a consistent set of geographies is used for the display, reporting and analysis of social, economic and environmental information. The common geographies should have a hierarchical design and support flexible reporting; covering a range from small geographic areas, such as a suburb or neighbourhood, through to larger areas, such as a natural resource management area or larger administrative jurisdictions. In order to connect people with place, these geographies should be designed to include approximately equal population numbers within each area across each hierarchy level. In the Australian context the common geography is the Australian Statistical Geography Standard (ASGS).

6.4 Statistical and Spatial Metadata Interoperability

Both the statistical and spatial data communities operate their own metadata capabilities. The statistical community use the Statistical Data and Metadata Exchange[1] (SDMX) and Data Documentation Initiative[2] (DDI) mechanisms. The spatial community has developed the ISO19115[3] metadata standard.

It is highly unlikely that either community would make any significant changes to fully adopt the other community's metadata structures, so an approach needs to be found that will enable discovery and access to the required statistical information for the geographic area of interest. This is one area of basic 'research and development' required for establishing the Statistical Spatial Framework.

6.5 Guidance Material

This component of the Framework is aimed at identifying guidance material relevant to the various activities that form part of the Framework. This information will highlight the wide range of legislative and operational issues that organisations need to be aware of when dealing with information about people and businesses; with the most significant issue being the privacy and confidentiality requirements for managing personal and commercial information. Other issues of relevance include: data quality, analysis, dissemination and visualisation.

The methodologies and approaches used by National Statistical Organisations, such as the ABS, offer examples of best practice that could be applied within other government organisations and commercial companies who need to address issues associated with geospatially enabled data. Practices applied in other organisations will also yield useful material to form part of the guidance material required for the Framework.

---

[1] http://sdmx.org/
[2] http://www.ddialliance.org/
[3] http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020

7. Conclusion

The ABS considers that the Statistical Spatial Framework will have applicability in any organisation wanting to understand people, society and the economy in the context of 'place'. At the international level the Framework could provide a common approach, enabling direct comparisons of many types of geospatially enabled statistical information, both within and between countries.

By standardising the process of integrating a range of socio-economic information within a location context, the Framework will empower people and organisations to:

- improve planning for regional economies and communities
- better target service delivery and business activities at the local-area level
- engage in more informed community level decision making
- incorporate population, social, economic and environment data into current geospatial and business analysis.

The Framework will also benefit the considerable efforts currently being made to bring a range of data together to better understand local-area level causes, impacts and responses to national and global concerns, such as natural disasters, climate change and sustainable development.

Acknowledgement

References

Australian Bureau of Statistics (2010), *Australian Statistical Geography Standard (ASGS)*, ABS Catalogue Number 1270.0, ABS website, viewed April 2013

Australian Bureau of Statistics (2013), *Connecting Information about People, Society and the Economy to a Location,* NSS website, viewed April 2013

ANZLIC – the Spatial Information Council, *National Address Management Framework*, ANZLIC website, viewed April 2013

United Nations (2013), 44th Session of the United Nations Statistics Commission, *Report of the Australian Bureau of Statistics on developing a statistical-geospatial framework*, E/CN/3/2013/2.