Distr. GENERAL
01 May 2013
WP.11
ENGLISH ONLY

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Work Session on Statistical Metadata
(Geneva, Switzerland 8-10 May 2013)
Topic (i): Metadata standards and models

Statistical and Spatial Frameworks. Standards and Data Infrastructure

Working Paper

Prepared by Al Hamilton, Australian Bureau of Statistics (ABS)

I. Introduction

1. This paper outlines how, in addition to official statistics, geospatial data has always been important to the information systems of a democratic society.

2. It explores the exponential growth in the availability of detailed geospatial data in recent years and the interest and ability of government, businesses and the public to make practical use of this data. This had led to a focus on Spatial Data Infrastructure (SDI), underpinned by data and metadata standards, at a national and transnational level.

3. There is an increasing focus by statistical agencies at the national and international level on seizing the opportunities to define and promote interoperability[1] between statistical and spatial data infrastructure.

4. A key element of promoting interoperability is better understanding, documenting and applying the relationships between relevant frameworks and standards related to statistical and spatial information (data and metadata). In recent months the High Level Group for Modernization Statistical Production and Services (HLG) has asked that participants in the METIS work program on statistical metadata contribute their professional expertise to assist with this work.

5. One purpose of this paper is to summarise context related to spatial frameworks, standards and data infrastructure for participants in METIS who may not be familiar with these topics. Readers may observe many parallels, and points of connection, with equivalent work related to statistical data and metadata.

[1] For the purposes of this paper, "interoperability" is defined in accordance with the European Interoperability Framework as "the ability of information and communication technology (ICT) systems and of the business processes they support to exchange data and to enable the sharing of information and knowledge."

6. It is hoped that broad familiarity with frameworks and standards related to spatial information, and recognition of analogies and intersections with frameworks and standards related to statistical information, will support METIS participants in contributing to efforts at the national and international level.

7. Section II of this paper explores why spatial data is important to the information systems of a democratic society and the exponential growth in availability and use of spatial data in recent years. It charts how this has led to a focus on Spatial Data Infrastructure (SDI) at national and transnational levels.

8. Section III briefly explores the development of data and metadata exchange standards for spatial[2] information. Much of this work predated, and then facilitated, the exponential growth in availability and use of spatial information. This can be seen as strong evidence in support of the proposition in the HLG Vision and Implementation Strategy that agreeing and applying "industry" frameworks and standards for statistical information will be a vital enabler for achieving greater interoperability and sharing of components at a practical level.

9. Section IV introduces the Global Geospatial Information Management (GGIM) Initiative which has been initiated in recent years by the UN Statistics Division (UNSD). Interests of the GGIM Committee of Experts include linking geospatial information to statistics.

10. Section V outlines the UNSC Program Review which was recommended by the second meeting of the GGIM Committee of Experts and initiated in October 2012. The Programme Review considered current geospatial capabilities and capacity within NSOs as well as proposed future directions.

11. Section VI outlines the recommendations arising from the Program Review in regard to developing an international statistical-geographical framework and the next steps that were endorsed by the UNSC.

12. Section VII provides a brief summary of the Australian Statistical Spatial Framework (SSF). While the SSF contains some elements which are specific to Australia (eg the specific statistical geography structure) the overall design, and many of the components, of the SFF may also be relevant for consideration when establishing an international framework (or national frameworks in other countries).

13. The conclusion poses a number of questions for statistical metadata experts about how we can position ourselves to contribute to the next steps at the international level and (where they exist) to developments at the national level.

II. Importance of, and growth in, Spatial Information

14. The first of the UN's fundamental principles of official statistics[3] is

[2] There is some debate as to whether the term "spatial information", "geospatial information" or "geographic information" should be used.

The eminent geographer Michael Goodchild argues geospatial is an "earth scale" subset of spatial - where the latter can range from the nano scale to intergalactic scale. Nevertheless, given the commonly accepted term "Spatial Data Infrastructure", "spatial" and "geospatial" are treated as interchangeable for the purposes of this paper.

In the report to UNSC in 2012 it was noted that the expert group on Global Geospatial Information Management had selected the term "geospatial" over "geographic" because the former is more comprehensive and more commonly used at present. This paper uses the former on that basis.

[3] http://unstats.un.org/unsd/methods/statorg/FP-english.htm

Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available…

15. While official statistics are indispensable to the information system of a democratic society, they are not the only form of information which that is relevant to a society and its public. One element of the test of practical utility is whether users can readily harness official statistics in conjunction with other types of information which are relevant to their needs.

16. Geospatial information has always been highly relevant in this context. For example,

    - It describes to the natural and built environment in which citizens live their lives and consume goods and services.
    - It represents a key dimension when considering equity of access to government (and other) services and when measuring efficiency and effectiveness of service delivery by governments.
    - It provides a meaningful point of common reference when "overlaying" and analysing data from multiple sources (eg official statistics, data about infrastructure and services, data from administrative datasets, data from social media)
    - It can provide a "my neighbourhood" perspective which is highly relevant to citizens and their families, as well as to local businesses.

17. Developments in recent decades have led to an explosion in the availability of detailed geospatial data. Such developments include

    - GPS (Global Positioning System) supporting fine grained geo-location. From a relatively expensive technology at first, positioning via GPS is now available via Smartphones and other common devices.
    - Cellular networks supporting determining of position (eg of a mobile phone) without needing GPS
    - RFID (Radio-frequency identification) making it easy to identify items and track them as they move through time and space (eg parcels, cars passing through a toll point)
    - Prevalence of software and databases which support automated geocoding (conversion to geographic co-ordinates) of both street addresses and IP (internet) addresses.

18. As described in Section III, significant standardisation of geospatial data and metadata for interchange purposes has been occurring since the 1990s. While this standardisation was led by expert practitioners working in (at that time) a relatively narrow and specialised field of information management, the "industry standards" they developed played an important role in enabling a more recent explosion in geospatial capabilities that are readily usable by members of the general public. Examples of these capabilities include

    - Google Maps (including Google Maps Navigation using GPS and Google Maps Transit using public transport schedules)
    - Google Earth
    - Open source, standards based, geospatial tools such as GeoTools and GeoServer[4]

19. The developments outlined above have led governments at the national and transnational level to focus in recent years on defining and developing Spatial Data Infrastructure (SDI)[5].

20. SDI can be defined as

    - a data infrastructure implementing a framework of geographic data, metadata, users and tools that are interactively connected in order to use spatial data in an efficient and flexible way, or

[4]http://en.wikipedia.org/wiki/List_of_geographic_information_systems_software
[5]http://en.wikipedia.org/wiki/Spatial_Data_Infrastructure

- the technology, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data, or
- a coordinated series of agreements on technology standards, institutional arrangements, and policies that enable the discovery and use of geospatial information by users and for purposes other than those it was created for

21. Examples at the transnational level include

    - The INSPIRE Directive, which aims to create a European Union SDI[6].
    - Development of UNSDI, led by the UN Geographic Information Working Group (UNGIWG)[7].

22. UN-GGIM (UN Initiative on Global Geospatial Information Management, which is discussed further in Section IV of this paper) currently catalogues 16 examples of SDI at a national level, including Australia, Canada, China, India, Korea, Switzerland and the US.[8]

23. The development, population and use of Spatial Data Infrastructure at a national and transnational level might be seen as currently outpacing that of "statistical data infrastructure". As outlined in Para 15, from a government and public perspective the aim is for the two sets of "data infrastructure" to work together.

24. Among other statistical agencies at a national and transnational level, the need to define and promote integration between spatial and statistical data infrastructure has been recognised, and acted upon, by

    - UN Statistical Commission (UNSC) and UN Statistics Division (UNSD) (See Sections IV to VI),
    - High Level Group for Modernization Statistical Production and Services (HLG) under the Conference of European Statisticians (see Section VI), and
    - Eurostat (which fulfils a prominent role in the INSPIRE Co-ordination Team)

III. Emergence of standards for geospatial information

25. Section II describes geospatial capabilities that are now commonly used by members of the general public and the development of SDI at national and transnational levels. A key enabler for these more recent developments was the establishment of commonly used standard for the exchange of geospatial data and metadata.

26. During the 1980s GIS (Geographic Information System) software became prevalent within the geospatial information industry.

27. The "geospatial information industry" in the 1980s and early 1990s was much less pervasive than it is today. In the early 1990s, GIS remained a relatively expensive capability which primarily targeted specialist users. Nevertheless, standardised formats for defining and exchanging geospatial information began to emerge within the industry. . At least one of the "defacto" standards which emerged at that time (shapefiles) remains important within the industry twenty years later.

28. The Open Geospatial Consortium (OGC) was founded in 1994. OGC has since grown to consortium of 475 companies, government agencies and universities participating in a consensus process to develop publicly available interface standards. The aim of OGC is to enable geoprocessing technologies to interoperate, or "plug and play".[9]

[6] http://inspire.jrc.ec.europa.eu/index.cfm/pageid/48
[7] http://www.ungiwg.org/documents
[8] http://ggim.un.org/sdi.html
[9] http://www.opengeospatial.org/ogc

29. OGC has standardised 50+ specifications. Some of the most commonly used of these specifications in the geospatial information industry include

    - GML (Geography Markup Language)
    - KML (Keyhole Markup Language) - used widely by Google
    - WMS (Web Map Service) and WFS (Web Feature Service) used by developers of web services using geospatial information

30. Other OGC specifications address topics such as Registry Services, Catalogue Services, Access Control, and Semantic Web from a geospatial information perspective.

31. It can be seen that many of the topics addressed by OGC in the context of geospatial information standards are also of interest to statistical agencies in regard to statistical information standards. The statistical community, however, lacks a standards consortium with so many members, and such a long history, as OGC.

32. The on-going focus of HLG since 2011 on defining frameworks and standards to underpin statistical modernization is a positive development in this regard. The HLG strategy includes establishing "industry standards" which fully and consistently support the needs producers of official standards and are based on existing standards from broader communities (eg SDMX and DDI). OGC have adopted similar approaches for a number of their standards (eg in regard to registry services and access control).

33. The growth of GIS during the 1980s also led to a focus on metadata. By the 1990s standardisation of metadata had begun at national levels, including through development of specifications by

    - FGDC (Federal Geographic Data Committee) in the US
    - ANZLIC (Australian and New Zealand Land Information Council - more recently referred to as the Spatial Information Council)

34. It was recognised international standardisation would be appropriate. Work to harmonise existing metadata standards began in 1999. ISO 19115 "Geographic Information - Metadata" was released in 2003.

35. ISO 19139 was released in 2007 to provide a standard XML representation for metadata that corresponded semantically with ISO 19115.

36. Many national and international geospatial information frameworks incorporate ISO 19115 or specific “profiles” on it (and, often, ISO 19139 as well). Such frameworks include

    - INSPIRE (EU)[10]
    - UNGIWG[11]
    - FGDC North American Profile of ISO 19115[12]
    - ANZLIC Metadata Profile[13]

IV. The Global Geospatial Information Management (GGIM) Initiative

37. The UN Geographic Information Working Group (UNGIWG) was formed in 2000 to address common geospatial issues - maps, boundaries, data exchange, standards - that affect the work of UN Organizations and Member States. The members of UNGIWG are departments, funds, programmes and specialised agencies within the "United Nations family".

[10] http://inspire.jrc.ec.europa.eu/documents/Metadata/INSPIRE_MD_IR_and_ISO_v1_2_20100616.pdf
[11] http://www.ungiwg.org/content/interoperable-services
[12] http://www.fgdc.gov/metadata/geospatial-metadata-standards
[13] http://spatial.gov.au/system/files/public/resources/anzlic/ANZLICmetadataProfile_v1-1_2007.pdf

38. The UN Initiative on Global Geospatial Information Management (GGIM) is a more recent development, which the UNSD was instrumental in establishing.

39. Partners in the GGIM Initiative include member states and international organizations outside the UN, as well as UN Agencies[14].

40. GGIM operates under the auspices of the UNSD and the UN Cartographic Section.

41. The National Statistical and Geographical Institute of Brazil presented a paper at UNSC in 2010 calling for a focus on GGIM. The report on the session records

The Statistical Commission...Recognized the importance of the integration of geographic and statistical information and the opportunities provided in that context by the swift development of information technology… [and]… Called upon all national statistical offices to actively participate, in partnership with relevant national authorities, in the further development of national geographic information capacity in the context of spatial data infrastructures, taking full advantage of information technology and focusing special attention on the area of improving statistical and geographic metadata compatibility…[and]… Requested the Statistics Division to convene a meeting of an international expert group, consisting of statisticians and geographical information specialists, to address global geographic information management issues;;

42. From the outset, therefore, GGIM was conceived as

    - requiring statistical and geographical information specialists to work together, and
    - placing special attention on improving statistical and geographical metadata compatibility

43. After preparatory work, the UN Committee of Experts on GGIM was created in July 2011 and met for the first time in October 2011.

44. The report of the Secretary-General to provide an update on GGIM to UNSC in 2012 noted

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…. It is therefore imperative that the professional statistical community, in partnership with the relevant national authorities in the geospatial information field, remain fully informed and engaged in the further development of national geospatial information capacities and in the corresponding establishment of national spatial data infrastructures.

45. At its second meeting in August 2012 the Committee of Experts agreed that one of nine areas for focus in the future was linking geospatial information to statistics.

46. As an initial step, it was suggested that a Programme Review by the UNSC would be helpful to support the development of a Statistical Geospatial Framework in National Statistical Systems. ABS volunteered to assist with conduct of the Program Review and to prepare a paper for consideration by the UNSC in 2013.

V. UNSC Programme Review

47. The objectives of the UNSC Programme Review were

    - To present a review of current geospatial capabilities and capacity within NSOs;

[14] http://ggim.un.org/partners.html

- To propose roles for NSOs in geospatial activities, with a particular focus on integrating statistical and geospatial information;
- To propose how geospatial activities could be further developed by NSOs within countries, and understand user needs driving particular geospatial data developments
- To explore how NSOs do, or should be doing, geocoding of their data; and
- To explore how to set standards that integrate data between the two communities.

48. A key source of information for the Programme Review was a Survey of Linking Geospatial Information to Statistics, circulated in October 2012.

49. While not all questions were relevant to all countries, the survey spanned around 50 questions that addressed topics such as

- National drives for geospatial information
- National Statistical Geospatial Information Capability, including questions about
    a. the relationship between the work of the NSO and the National Geospatial Information Authority (NGIA)
    b. possible leadership roles assumed by the NSO, such as geospatially enabling national administrative and statistical data
    c. statistical geography and what geospatial attributes (if any) were attached to unit record and aggregate data
- Current and planned future integration between statistical and geospatial information capabilities within the country (including possible needs for standards and frameworks to enable improved integration)
- Benefits for the NSO and the country from linking geospatial information to statistics
- Successful case studies of linking geospatial and statistical information

50. 53 NSOs responded to the survey, creating a valuable dataset recording current practices and future plans.

51. A detailed analysis of survey responses was prepared by the ABS as a background document for the 2013 session of the UNSC[15].

52. Section II of the main report from the ABS, which was presented to UNSC for discussion and decision, provides a higher level summary of findings[16].

VI. Developing an international statistical-geospatial framework

53. The overall focus of the main report from the ABS was developing a statistical-geospatial framework including

- The need for linking socioeconomic information to a location
- The current situation (summarising the findings of the survey)
- The future information agenda (such as requirements beyond the Millennium Development Goals and to inform Sustainable Development).
- Proposed future directions

54. Section IV of the main report included a number of specific recommendations. These were (with selective abbreviation applied)

- An international conference be convened to identify and address common issues relating to linking socioeconomic information to a location, including developing best practice principles;
- Linkages between relevant statistical and geospatial organizations be formalized, building on the efforts of the UN Committee of Experts on GGIM and working with other relevant international entities, including HLG.

[15] http://unstats.un.org/unsd/statcom/doc13/BG-GGIM.pdf
[16] http://unstats.un.org/unsd/statcom/doc13/2013-2-ProgReview-E.pdf

- The approach used in Australia through the Statistical Spatial Framework be examined as a possible methodology to guide a common global approach to linking socioeconomic information to a location.
- A group of experts be established at an international level to further the development of a common approach to linking socioeconomic information to a location.
- In developing national statistics plans, countries be encouraged to consider the possibilities for linking statistical and spatial information, consistent with their development priorities.
- As national statistics offices undertake information management infrastructure transformation activities, consideration be given to adding geospatial capability, including the geocoding of addresses

55. In response to the report, the UNSC[17]

- Recognized the importance of the integration of geospatial information and statistics in supporting social, economic and environmental policy decision-making, including at the subnational level;
- Strongly supported the linking of social, economic and environmental data to time and location attributes in order to enrich and maximize the potential of statistical information, while noting the need to provide technical assistance to countries, developing countries in particular, in the early stages of the integration process;
- Welcomed the proposal to organize an international conference as a way of reaching out and developing best practices, bringing together both statistical and geospatial professional communities, while bearing in mind the critical link with the informatics community, especially in the context of the current discussion on big data;
- Also welcomed the proposal to develop an international statistical geospatial framework, taking into account existing national and international efforts;
- Requested the UNSD to establish an expert group composed of representatives of both statistical and geospatial communities to carry out work on developing a statistical-spatial framework as a global standard for the integration of statistical and geospatial information, and thanked the ABS for offering to provide continued leadership;

56. It is currently anticipated that the international conference will be held alongside another major related conference or meeting during the second half of 2013 or the beginning of 2014 and that the statistical-spatial expert group may be formed in the lead up to the conference.

57. HLG recognises it has a contribution to make when developing a statistical-geospatial framework at the international level. Such a framework will form an important guide for a number of aspects of modernising statistical production and services.

58. Following the GSIM Development Project in 2012, HLG has defined two major projects for 2013.

- Frameworks and Standards for Statistical Modernization Project
- Common Statistical Production Architecture Project ("Plug and Play")

59. Within the Frameworks and Standards Project[18], Work Package 6 is intended to provide an initial assessment of the role of geo-spatial standards in the modernisation of official statistics, including how they may relate to the GSBPM and the GSIM.

60. Examples of possible areas for review include

- How should relevant attribute components within unit and dimensional data structures used in statistics be denoted as geospatial in nature (eg if they contain geographical co-ordinates or codes for geographic areas) and linked to relevant geospatial metadata (eg spatial representation information, reference system information)?

[17] The list is a summary of selected points recorded under 44/101 in the Report on the forty-fourth session of the UNSC

[18] http://www1.unece.org/stat/platform/download/attachments/58492100/Standards+project+outline.docx

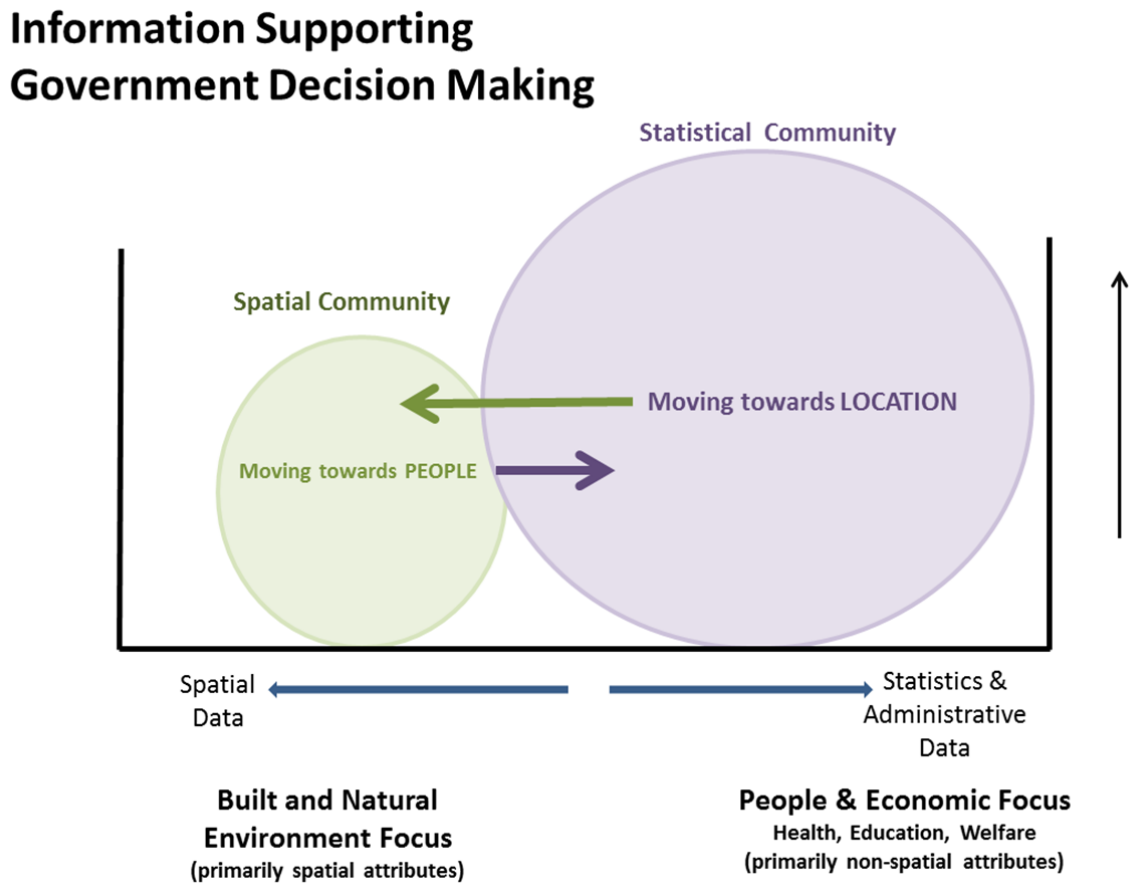
- Both statistical and geospatial metadata include data quality information. Quality in the latter case typically refers to precision of spatial positioning. How can we manage the two types of quality information, making the right information available to the right user?
- Both statistical and geospatial metadata include units of measure. How do we rationalise/integrate the information?
- ISO 19115 has a concept of data lineage which is similar to the concept of provenance for statistical data. How do we rationalise/integrate information?

61. Work Package 6 may be expected to make an important contribution to the broader international work on establishing an international statistical-geospatial framework.

VII. What might an international statistical-geospatial framework entail?

62. The Australian Statistical Spatial Framework (SSF) was presented to the UNSC as an example at the national level of a type of framework which might also be relevant at an international level.

63. A key driver for establishing the SFF is the growing intersection between the way the Spatial Community and the Statistical Community inform decision making by government.



64. The concept of location or 'place' is now a key driver for the ABS and other organisations collecting, compiling, analysing and disseminating socio-economic statistical information

65. SSF is aimed at providing a consistent and common spatial approach for all providers of socio-economic information. Using a common approach will greatly simplify the process of linking socio-economic data sets to help better understand a wide range of complex issues, improving the ability of government and the community to make more informed decisions.

66. SSF is essentially a bridge - a bridge between the statistical and spatial communities and the systems in which they operate. The common element in this bridge is geography. Geography draws the socio-economic into the spatial community's environment, and makes it available for use within that environment.

67. Relevance and fitness for purpose of SSF as a framework needs to be considered at four levels

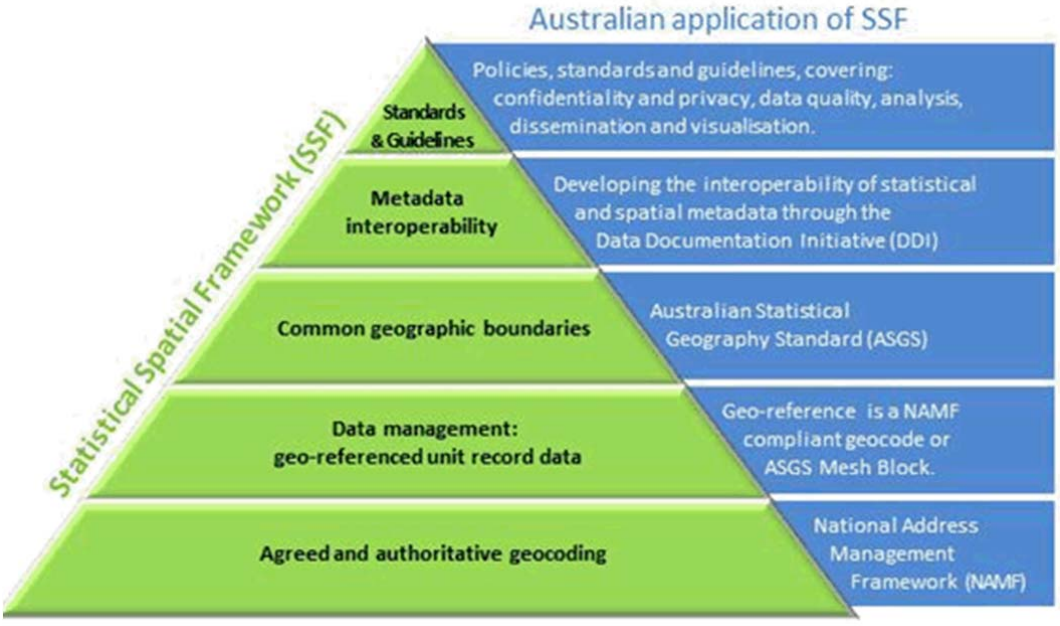
    - ABS
    - Australian Government
    - Nationally (including state and local government, businesses, community)
    - Internationally (eg alignment with international statistical and spatial standards)

68. The SSF is built on existing standards and approaches including the Australian Statistical Geography Standard (ASGS), developed by the ABS. It also includes the National Address Management Framework (NAMF), a consistent method of applying a location (latitude and longitude) to an address. NAMF has been endorsed by all Australian jurisdictions.

69. While the specific standard for statistical geography and specific framework for coding addresses will vary from country to country, these would be important components for any nation's statistical-spatial framework.

70. The five main elements of the SSF are

    - An agreed approach to determining geo-position based on an address using NAMF
    - Storing the position data together with the unit level records in organisations' data management systems (understanding that the unit level data is unlikely to be released).
    - Using the ASGS to provide the hierarchy of consistent geographic boundaries.
    - Linking the geospatial and statistical metadata capabilities to support geographic searches of statistical data.
    - Having a range of best practise and possibly legislative protocols for supporting traditional statistical issues such as privacy and confidentiality.



71. The intent and design of the SSF was discussed with a number of key stakeholders during 2012. These included

    - Australian Government agencies
    - State jurisdictions

- ANZLIC (now known as the Spatial Information Council)

72. The feedback was highly encouraging, with a number of agencies beyond the ABS being eager to adopt the framework.

73. By standardising the process of integrating a range of socio-economic information within a location context, the SSF is expected to enable

    - ·Improved planning for regional economies and communities;
    - ·Targeted service delivery at the small area level; and
    - ·Community level decision making

74. In addition, the Framework will support the considerable efforts currently being made to bring a range of data together to better understand the causes, impacts and responses at the local level to national and global concerns such as climate change and sustainable development.

75. In summary, the SSF is seen as supporting environmental, social and economic intelligence in the twenty-first century.

VIII. Conclusion

76. Spatial data is increasingly prevalent and increasingly used by governments and the community.

77. There is widespread recognition, including by the UN Statistical Commission, of the importance and value of the integration of geospatial information and statistics in supporting social, economic and environmental policy decision-making

78. Developing, agreeing and applying statistical spatial frameworks can facilitate this integration at the national and international levels.

79. UNSC, UNSD and HLG are committed to taking practical steps to establish such frameworks.

80. The expertise and experience of specialists in statistical information management, including statistical metadata, is expected to make a vital contribution to defining frameworks which serve as the "bridge" between the spatial and statistical communities.

81. Participants in the METIS Work Session may like to consider questions such as

    - What additional information and/or actions might help them better to contribute to this work at the national and/or international level?
    - How might contributions from the statistical metadata community best be progressed?
    - What collaboration arrangements would be best across the statistical metadata community and beyond that community (eg with geospatial data and metadata experts)?