United Nations E/CONF.101/55

Economic and Social Council

Distr.: General

21 May 2012

Original: English

Tenth United Nations Conference on the Standardization of Geographical Names
New York, 31 July – 9 August 2012
Item 11(d) of the provisional agenda*
Toponymic data files and Gazetteers
Data standards and interoperability

UNSDI Gazetteer Framework for Social Protection in Indonesia

Submitted by Australia**

* E/CONF.101/1.

** Prepared by: Paul Box, Laura Kostanski and Rob Atkinson (Australia), Commonwealth Scientific and Industrial Research Organisation(CSIRO), with assistance from: Suha Ulgen, Co-chair, UN Geographic Information Working Group (UNGIWG) & Head, Centre of Excellence for UN Spatial Data Infrastructure (CoE4UNSDI); -Pier-Giorgio Zaccheddu, Federal Agency for Cartography and Geodesy (BKG)

Summary

The Gazetteer Framework Project aims to improve integration of, access to and use of data held in different government systems- through the development of a framework to manage gazetteers (directories of place names or identifiers with associated location information).

The Gazetteer Framework is envisaged to provide a mechanism to link spatial data to support a range of development, humanitarian and environmental activities at different scales. The disparate systems which require integration to support operational outcomes typically employ different approaches to the use of gazetteers. Consequently, it is time-consuming and expensive to discover, access, interpret, transform and integrate information from different sources referenced using different gazetteers. Furthermore, when data changes the process must be repeated.

A particular challenge in this process is the interpretation and matching of spatial identifiers across multiple datasets i.e. determining whether different data sources are using the same or different place names and associated locations/boundariesto reference information. Therefore, to enable multi-sector analysis and the development of coherent, complimentary, multi-sectoral responses, it is critical that the spatial dimension of information held by agencies working in different sectors can be easily understood.

This project aims to re-engineer gazetteer data management, publishing and access mechanisms through the use of “game changers” in computing and networking (ICT) infrastructure. These include the use of LinkedData mechanismsand “cloud computing”to provide a ubiquitous, Web accessible and flexible infrastructure for to enable the resolving of spatial location identifiers providing links to definitions, spatial representations and related data.

The project is being funded by the Australian Agency for International Development (AusAID) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO)and is initially focused on supporting Social Protection outcomes in Indonesia. The projecthas been jointly developed by CSIRO and the Office of the Assistant Secretary-General and Chief Information Technology Officer (ASG/CITO) of the UN Secretariat and is adeliverable of the Centre of Excellence for UN Spatial Data Infrastructure (CoE4UNSDI). The CoE4UNSDIaims to support development of standards and best practice methodologies for provisioning core datasets used within the UN System by strengthening capacities to create, maintain and publish spatial information. Within this context the UN Gazetteer Framework aims to provide standardised and easily exploitable means to reference geographic locations to support applications of the UNSDI.

As part of the UNSDI Standards and Best Practices activity the project intends to collaborate with relevant expert groups including UNGEGN. UNGEGN members are encourage to actively participate in this project and plan to avail themselves of the infrastructure being developed to support their ongoing activities. Demonstrations of the Gazetteer Framework will be provided during the UNCSGN and UNGEGN members are invited to discuss collaboration potentials with the project team members.

Overview

Global and local shocks such as the financial crisis, tsunamis, earthquakes, and crop failures have had significant impact on the livelihood of people living in or close to poverty. In response, there has been a renewed focus on social protection in developing nations which aims to buffer vulnerable populations from shocks through social assistance and other measures. The UN Global Pulse initiative was created to assist in monitoring vulnerable populations using new sources of information (such as social media) and real time analytics, to gain new insights in changes in human wellbeing.

The UNSDI Gazetteer Framework for Social Protection in Indonesia (the Gazetteer Framework) Project will support the Global Pulse initiative by increasingaccess to and integration of spatial and other information for social protection, vulnerability analysis and poverty reduction, using gazetteers.The project will deliver a flexible standards-based approach to gazetteer infrastructure that can be used in a global context, across any application domain that needs to integrate geographic data from multiple sources.

To assess and monitor vulnerable populations, accurate, up-to-date and fine-grained information about communities and their exposure to hazard is required, particularly given the spatial variability in environmental characteristics of land and the socio-economic characteristics of its inhabitants. Effective coordination of response requires a common understanding of the location, condition and needs of the target population together with a shared situational awareness of interventions by government and other actors. Therefore, to enable multi-sector analysis and the development of coherent, complimentary, multi-sectoral responses, it is critical that information held by agencies working in different sectors, can be readily accessed and integrated.

However, government and other systems that contain the required information use different approaches to referencing information based upon the use of gazetteers or directories of place names with associated location information. Gazetteers are used to spatially reference or ‘geo-code’ information holdings and there are typically many gazetteers in use, including gazetteers sourced from topographic maps, locality gazetteers developed through census activities, administrative area gazetteers as well as an increasing number of informal, crowd-sourced gazetteers. Consequently, it is time-consuming and expensive to discover, access, interpret, transform and integrate information from different sources referenced using different gazetteers. Furthermore, when data changes the process must be repeated. The gazetteer framework developed through this project is an attempt to addressthese challenges through improvedgazetteer management, publication and accessmechanisms.

Gazetteer Definition

A *Gazetteer* is commonly considered to be a list of place names and location details (i.e. coordinates, address), classified by feature type (i.e. mountain, building, river). A broader view is that anyset of identifiers for locations which may not have names (e.g. postcodes, or asset numbers) may also be used as a gazetteer. This project takes this broader view and thus,a gazetteer is considered to be:any geospatial dataset defining a set of features that may be used as a location reference by other data sets. *For example - an electronic index of schools with their coordinates, or a dataset of airports with their names and location details.*

Often, multiple datasets will contain reference to the same features. For instance, a government health department and a non-government health organisation might both have datasets listing regional

hospitals. Or, a national census bureau and a national planning agency might both have datasets on populated places and administrative boundaries.

Theproject aims to create a system for registering and linking multiple disparate gazetteers through the cross-referencing of placenames and identifiers. This approach addresses the reality that there are different gazetteers in use and that attempting to develop a single harmonised data set covering all possible feature types is not feasible nor desirable. The linking of gazetteers will enable government and non-government agencies to reliably publish, access and share information with a reduction in confusion caused by ambiguous names, mismatched boundaries, missing data, misattributed coordinates and incorrect information.

In this project a gazetteer is simply a data set used to understand relevant location references, accessible by a simplified and flexible set of mechanisms through the UNSDI Gazetteer Framework infrastructure.

Multiple overlapping gazetteers in use

Although different sectors or communities may utilise different gazetteers, reflecting the needs and realities of the sector, two gazetteers are commonly used across multiple sectors: 1) populated places showing the location and name of settlements, and 2) administrative units, which reflect the hierarchical, territorial arrangement of government administration. In many cases, these gazetteers can be out of date (as they are based on expensive and thus infrequent field data collection efforts) or are inaccessible due to the sensitivity around spatial information access.

Consequently, users may opt to create their own gazetteers, or end up using different versions of gazetteers, which diverge over time. For these reasons it is extremely time-consuming and unreliable to interpret and integrate information from different sources based on different gazetteers. Figure 1 below, illustrates the current situation and the impact of the proposed solution on the relative amount of time and effort spent on discovery and preparation as opposed to use of information.

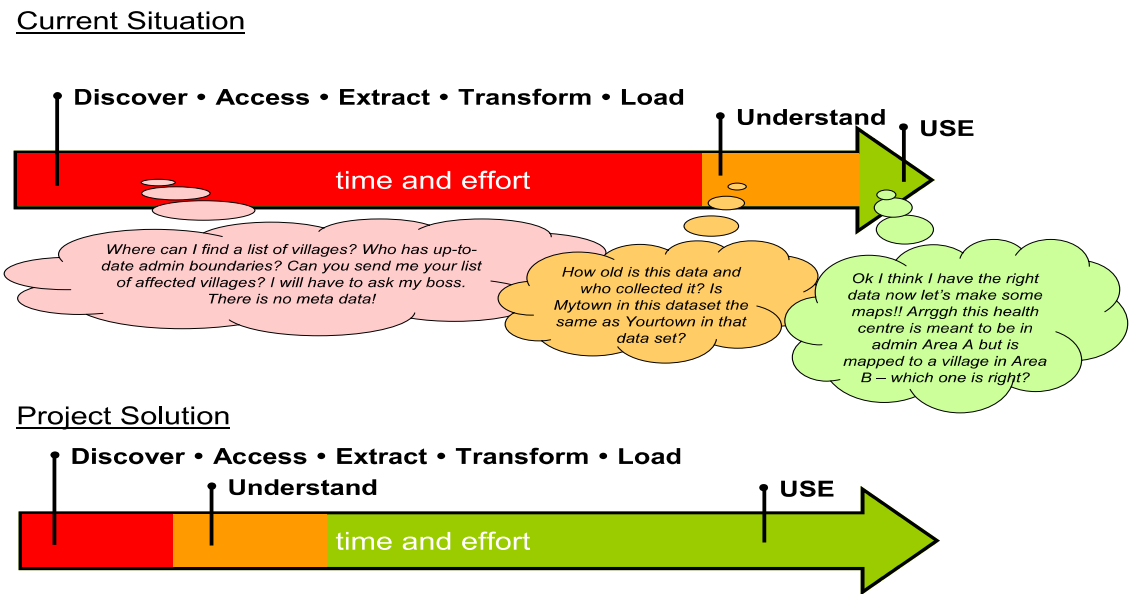

Figure 1Time and effort spent in discovery, access, preparation and use of information (current and proposed future situation)

A great deal of effort is duplicated as each agency or person attempting to integrate data, must solve the same set of challenges. A key challenge is the interpretation and matching of spatial identifiers across multiple datasets i.e. determining whether different data sources are referencing the same or different places.Furthermore, having solved these challenges (i.e. having linked datasets) to conduct analysis or to produce situational awareness maps, the link between these datasets is not easily shared with others, decays over time as changes to the related data sets occur, and is subsequently lost. This presents a significant wasted effort and associated costs.

The role of crowd-sourced information and new technologies

Increasingly, producers and users of formal information are looking to volunteered and crowd sourced information provided by citizens, to provide additional and very rapidly produced insights into situation in people's lives. It is hoped that this volunteered information will enable better informed and more agile policy responses and interventions.

The recent emergence of crowd-sourced gazetteers has added to the challenges of integration of data referenced using gazetteers. In a range of application areas, crowd-sourced (volunteer contributed) information is playing an increasingly important role. In disaster response for example, information contribution has taken the form of remote volunteered effort as in the assessments of earthquake damage in Haiti undertaken by using OpenStreetMap or from responders and affected populations on the ground sending messages (SMS) using mobile phones. This information consolidated in platforms such as Ushahidi and presented in innovative new ways provides fast accurate data to responders on the ground. The challenges associated with integrating traditional, authoritative gazetteers with emerging informal crowd-sourced gazetteers must be addressed to maximise the benefit of these important information sources.

Crowd-sourced data is usually built upon available gazetteers. The provenance of these gazetteers is not managed in a way that allows new information about new or changed places to be reliably fed back to the original sources. Thus the considerable volunteered effort to supplement gazetteers is not maximised in terms of improving existing gazetteers. A range of new technologies most notably, social network platforms such as Twitter and Facebook have emerged that enable global discussion, collaboration, and mobilisation of the crowd to solve large scale problems. It is possible to mine these platforms to analyse the content of what is being discussed.

The Global Pulse initiative intends to leverage these technologies and mine the rich available seams of information that they provide in order to get deep and timely insights into issues that people are concerned about. The key to maximising the benefits of these powerful new information sources will be to integrate and analyse them in parallel with the more formally collected and managed sources of information curated by government and other agencies.

Project Goals

The goal of the project is to achieve fundamental, systemic improvement in information integration that enables more effective and cost-efficient service delivery.To realise this goal, the project will develop a

gazetteer framework that supplements and supports national efforts to build, maintain and provide access to spatial data. This framework will provide a common mechanism for the use of gazetteers that will enable them to be embedded within systems to reference socio-economic and environmental data, in such a way that enables this information to be more easily integrated. The framework will also leverage the power of volunteered geographic information and provide a means of reconciling volunteered information with more formal government and other data sources, consistent with the overall goals of the Global Pulse initiative.

The Gazetteer Framework

The gazetteer frameworksupplements and supports national efforts to build, maintain and provide access to spatial data. The framework will provide a common mechanism for the publication of, and access to gazetteers that will enable them to be embedded within systems to reference socio-economic, environmental and other data. In turn, this will make information from different systems more readily accessible, and more easyto integrate.

The gazetteer framework will be used as a critical spatial referencing framework in themonitoring platform and tools being developed by Global Pulse. The project will establish the enabling gazetteer framework itself, working with Global Pulse to determine priority gazetteers usage. The framework will enable gazetteers, together with the individual place names used in each gazetteer, to be registered. Thus, the gazetteer framework acts as a register of gazetteers, the systems that use these gazetteers and the individual place names within each gazetteer. Specifically, the framework will provide mechanisms for:

- registration of authoritative source gazetteers;
- harvesting, indexingand integration of registered gazetteersinto a common gazetteer, including the creation of cross-walks between gazetteers terms;
- registration of information systems that utilise gazetteers, thus enabling the creation of a social network graph of gazetteer usage;
- users to search for place names across all registered gazetteers, explore relationships between place names in different gazetteers and unambiguously reference and thus use individual place names i.e. ‘cite’ or use place names as ‘spatial bookmarks’; and
- users to add their own place names to supplement registered gazetteers and to tag existing place names in gazetteers to report deficiencies to relevant authorities and thus improve the data quality.

The framework is intended to act as persistent infrastructure which supplements existing systems, making them easier to access, and enabling common usage of improving resources rather than fragmentation of efforts attempting to maintain duplicative and increasingly divergent gazetteers.

The framework addresses concerns at two distinct, yet related, scales of operation. At the national scale, focusing on Indonesia, the project aims to understand and map the methods for managing, integrating and accessing Indonesian gazetteers. At a global scale it is theorised that the same mechanism could be used to improve the management of and access to identified gazetteers as a critical element of the UNSDI. These scales of activity are intrinsically related as a UNSDI is in effect an

aggregation of national SDI data offerings together with global datasets that are created and managed by the UN System organisations.

The project is underpinned by an innovative methodology and toolset for system design that could potentially be applied to numerous information integration challenges that will need to be addressed in order to build the UNSDI. Through this approach, the project has the potential to address universal infrastructure governance issues – a necessary precursor for internationally-sustainable infrastructure growth.

Collaboration

The Gazetteer Framework Project has been jointly developed by CSIRO and the Office of the Assistant Secretary-General and Chief Information Technology Officer (ASG/CITO) of the UN Secretariat and is a deliverable of the Centre of Excellence for UN Spatial Data Infrastructure (CoE4UNSDI). It has been designed to provide a basis for ongoing development and evolution of the UNSDI. Specifically, it aims to support development of standards and best practice methodologies for provisioning core datasets used within the UN System by strengthening capacities to create, maintain and publish spatial information.

The project is being implemented in collaboration with BadanInformasiGeospasial (BIG) the Indonesian National Agency for Geospatial Information. BIG is responsible for the national gazetteer and leads Indonesian Government efforts to build the Indonesian Spatial Data Infrastructure (InaSDI). The capabilities developed by the project will be integrated into InaSDI leading to improved integration and use of gazetteers as key index mechanism for spatial data delivered through the SDI. This approach will be used to build a broaderUNSDI capability, with a set of components and approaches being developed thatcan be adopted by both UN agencies and Member states to augmenttheir existing SDI activities, and in turn make the data these organisations manage more easily integrated with data held by other organisations.

The project is engaging with a range of international activities to co-develop appropriate technical standards to future-proof the framework. This includes collaboration with the OpenGeospatial Consortium (OGC), ISO, UK Location Strategy and the Joint Research Centre of the European Union.

Research agenda

From a research perspective the project is focused on a single broad question: “How might shared infrastructure be used to improve interoperability of information referenced using spatial identifiers, to provide new linking mechanisms which enable timely multi sectoral/cross domain decision making?” The key aspects of this broad question that are the focus of research are as follows:

Interoperability - “Everything happens somewhere” and thus access and exploitation of spatial identifiers utilising semantic tools provides a powerful mechanism for integration, analysis and presentation of a wide range of information. The Gazetteer Framework project is focused on applications of interoperable spatial systems more commonly referred to as Spatial Data Infrastructures (SDI). The project will identify preferred methods for cross-referencing and linking components of SDIs as they relate to identifiers contained in gazetteers.

Information referenced using spatial identifiers – spatial referencing frameworks are a logical partitioning of space into discrete related units (administrative units, hierarchies of settlements). The Gazetteer Framework project will explore existing methods for managing, integrating and using gazetteers as spatial reference frameworks. The intention is to identify potential approaches to improve access to, and integration of, information referenced by gazetteers held in multiple systems.

Multi-sectoral and cross-domain decision-making - Traditional vertical information flows are tending towards horizontal and diagonal- resulting in increased cross-sector information data access and information matching requirements. As this work becomes more prevalent, there is increased demand for rapid access to, and potential for integrating with, information from a growing number of systems (for example matching statistical information held by a national census bureau with the administrative regional boundaries maintained by a national mapping agency). Currently, each data information system is typically designed for a specific purpose and therefore can have a limited ability to exchange information with other systems (for example the census information is contained within a dataset maintained within a government statistics department data system which is unlinked and unrelated to the government mapping department data system). Today, it is recognized that systems can be interconnected in ecosystems often referred to as 'systems of systems' (SoS) (Cantot and Luzeaux, 2011). SoS are generally theorised as solutions for achieving interoperability and superior coordination between heterogeneous systems (Jamshidi, 2008). However, building a system that can easily exchange information between its component parts i.e. is interoperable, presents a range of challenges. These challenges are a key focus of the Gazetteer Framework project.

In addition to research into the role of gazetteers as interoperability mechanism, the project is also focused on 'core gazetteer research' to improve development, management and use of gazetteers.This aspect of the research will concentrate on toponymy and the identification of places through spatial data referencing systems. The research will explore processes improvement for decentralising and federating the management of gazetteers and the development of optimised methods for data provision based on user requirements. In addition, the research will explore approaches to supporting crowd-sourcing techniques and the integration of formal and informal information into SDIs.

Progress to date

In the initial, one yearpilot phase of the project which is nearing completion,a prototype gazetteer framework was designed, tested and presented via a demonstrator application. The demonstrator has beenreviewed with a broad range of stakeholders: inIndonesia withGovernment of Indonesia stakeholders including BIG and, and with a range of UN and other stakeholders at a global scale. This demonstrator has enabled CSIRO to present and validate proposed approaches to managing, linking and improving the re-use of gazetteers.

In addition to the Indonesian national gazetteer that has been the focus of the project, the Australian national gazetteer and the UNGEGN global gazetteers have also been harvested into the framework as test data sets. The gazetteer framework has received strong support in Australia with the Committee for Geographic Names for Australasia (CGNA) and Office of Spatial Policy endorsing the initiative. Current efforts to upgrade theAustralian national gazetteerare being aligned with theUNSDI Gazetteer Framework.

Demonstration & Further Information

Technical documentation and demonstration will be provided at a workshop during the UNCSGN. Further information can be located at www.csiro.au/gazetteer

Conclusions

The project attempts to improve the quality and timeliness of information critical for targeting and monitoring long-term poverty reduction and social protection through the development of a gazetteer framework, though the framework is equally relevant to a range of other applications. In essence, it functions as a 'Rosetta stone',using gazetteers to link into multiple systems and the references to place names that these systems use. This approach is also intended to assist in bridgingcrowd-sourced and formal data sources so their respective capabilities are complimentary.

As part of the UNSDI Standards and Best Practices activity the project intends to collaborate withrelevant expert groups including UNGEGN. UNGEGN members are encourage to actively participatein this project and plan to avail themselves of the infrastructure being developed to support their ongoing activities.

Glossary

Gazetteer – classically defined asa dataset containing a list of place names. Classification as a gazetteer is determined when another dataset uses the place name list as a referencing tool. A broad definition of place names is used for the research, and consistent with *ISO 19112 (Spatial Referencing by Identifier*)(2003) the Gazetteer Framework project generally refers to '**identifiers**'. Gazetteers can be official/ unofficial, formal/informal, authoritative/non-authoritative. The definition of these types is considered to be a complex issue to be resolved during the process of the project research and further discussion is provided in supplementary UNGEGN paper- *Four Faces of Toponymic Gazetteers*-submitted by Australia.

Common Gazetteer – an integrated index of identifiers that have been harvested from individual **source gazetteers**. A common gazetteer comprises a list of identifiers with associated properties (identity, type, location, source gazetteer metadata) formatted using a common information structure. For example, the Gazetteer of Australia is comprised of toponymic source gazetteers from Australian jurisdictional naming authorities- the data of which is reformatted according to the requirements of the national gazetteer information model.

Source gazetteer – dataset containing a list of identifiers that may be used as spatial references for other geographical information. They can be reference datasets or those used for particular application processes and their information structure can take many forms. They are a recognised component of a common gazetteer and identifiable by particular governance structures and management cycles. For example- a dataset of populated places comprising name, unique reference code and position details. A source gazetteer may or may not be accessed directly using a variety of interfaces (such as WFS).

Application/Project gazetteer – a type of source gazetteer that has been developed for a specific project or application. For example, a list of Internally Displaced Persons (IDP) camps including name, unique reference code and location. This becomes a **source gazetteer** if another stakeholder subscribes to it and should be included in the **common gazetteer** (act as a source gazetteer) if another dataset uses it for reference.

Identifier- a shareable reference to a single geographical object. Examples include place names, postcodes, asset numbers etc.

References

2003. ISO 19112:2003 Geographic information -- Spatial referencing by geographic identifiers. *In:* ORGANISATION, I. S. (ed.). www.iso.org

CANTOT, P. & LUZEAUX, D. 2011. *Simulation and modelling of systems of systems,* London, ISTE.

JAMSHIDI, M. 2008. *Systems of Systems Engineering: Principles and Applications,* CRC.