stepSA
SPATIAL TEMPORAL EVIDENCE FOR PLANNING **SOUTH AFRICA**
STEPSA.ORG

Policy Note 2

Spatial data infrastructures (SDIs) and observatories for spatial analysis

Objective of the note

This policy note highlights some issues concerning spatial data infrastructures (SDIs) and similar systems in South Africa, which aim at providing integrated, harmonized, quality geospatial data sets readily, quickly and affordably to users. SDIs are particularly useful for supporting evidence-based planning.

Please note that there is a variety of terms used for geospatial data, but they are all essentially synonyms: spatial, spatially-referenced, geographical, geographically-referenced, geo-referenced or geo, data or information. The term *geospatial* shall be used here.

Background to geospatial data

Geospatial data have many applications in many fields, particularly in planning. With the ready availability of virtual globes, such as Google Earth; consumer GNSS[1] receivers, embedded in the likes of vehicle navigation systems and mobile telephones; online repositories of user-generated geospatial data (also known as volunteered geographical information, or VGI), such as OpenStreetMap and Tracks4Africa; geospatial social media or networking services (geo-social services), such as Foursquare; geocoding of articles and photographs, such as in Wikipedia or Panoramio; and other location-based services, there is now a much greater awareness amongst the lay public and professionals of the power and utility of geospatial data.

As a consequence, there is a growing demand for geospatial data. Technological advances such as total stations and other surveying equipment, and the proliferation of sensors such as satellite and aerial imagers, together with the many amateurs now contributing geospatial data and helpful legislation (eg: South Africa's Promotion of Access to Information Act [2000]), make massive volumes of digital geospatial data readily available. Unfortunately, geospatial data are inherently complicated because the Earth is an irregular oblate spheroid, requiring a variety of coordinate reference systems to project the data[2], and not a three-dimensional Cartesian space, which is how some try to treat geospatial data. Further, abstract concepts within geospatial data have to be made concrete to be rendered to data structures and code within any digital system, and there are complexities concerning scale, multiple representations,

[1] Global navigation satellite systems, including the USA's NAVSTAR or GPS (global positioning system), Russia's GLONASS (global navigation satellite system), China's Beidou, etc.
[2] All projections are a compromise between preserving direction, shape, area, distance and/or shortest route, and can have different datums (models of the Earth).

CSIR our future through science
science & technology Department: Science and Technology REPUBLIC OF SOUTH AFRICA
HSRC Human Sciences Research Council

maintenance, data quality (currency or timeliness, spatial accuracy, attribute accuracy, resolution, consistency, completeness, etc), provenance, metadata, classification vs. attributes, and topological and other relationships between features (the phenomena in the data set). Any set of digital data is always only just one of many possible 'views' of the world – they cannot be an exact duplication of the world. For any data set, some things are approximated, some things are simplified and some things are ignored. Hence, there can never be perfect, complete and correct data.

So, even though vast quantities of geospatial data are readily available, even for free or very cheaply, they are not necessarily easily discoverable or suitable for use by many users. Nor is it easy to assess the suitability of the data for one's application (i.e.: the fitness for use)

The result is that various initiatives have and are being taken to "facilitate the capture, management, maintenance, integration, distribution and use of spatial information'' [South Africa 2003]. Unfortunately, these initiatives are not necessarily integrated and/or complementary, but can be competing with one another. Part of the problem is that because of the complexities highlighted above, it takes a long time to establish an SDI, while many are looking for a quick fix. Confusingly, these initiatives use different terms for what are essentially the same things: SDI, spatial observatory, geoportal, geospatial data clearinghouse, geospatial platform, digital earth, etc.

Background to spatial data infrastructures

One of the distinguishing characteristics of the use of geospatial data is that the same, common, base geospatial data sets[3] are used by many different users for many diverse applications. Hence, there is a growing need to share and organise geospatial data across different disciplines and organisations, which has resulted in the development and implementation of spatial data infrastructures (SDIs) and of the theory and notions behind them. An SDI is an evolving concept about facilitating and coordinating the exchange and sharing of geospatial data and services between stakeholders from different levels in the geospatial data community [Hjelmager *et al* 2008]. An SDI is more than just the technology of a geographical information system (GIS): it is generally considered to be the collection of technologies, policies and institutional arrangements that facilitates the availability of, and access to, geospatial data. It provides a basis for geospatial data discovery, evaluation and application for a variety of users and providers [Nebert 2004]. An SDI should facilitate mashups, combining content from multiple sources, even multimedia.

Essentially, base geospatial data should be considered separately from the value-add geospatial data of users, as the latter are often not for sharing, and are often for limited use, transient, parochial and/or even poorly documented. There are two uses for base data that are quite different: the key data providing structure and connections for other data sets, and the important and widely used data sets [Schwabe & Govender 2012], though many base data sets are used for both. A major problem with the first type of use is that of *incremental updating and versioning*: when a base data set is updated and features are deleted, merged or moved to reflect changes in the real world or to correct errors in the data, these modifications can remove or alter the geocoding of a user's value-add geospatial data [Cooper & Peled 2001].

South African Spatial Data Infrastructure

The Spatial Data Infrastructure Act [2003] established the Committee for Spatial Information (CSI) and tasked it with setting up the South African Spatial Data Infrastructure (SASDI). For various reasons, it only came into effect in 2010,

[3] Also known as reference, framework, primary, fundamental, core and foundation data sets [Schwabe & Govender 2012].

when the CSI was appointed and first met. To date, the CSI developed criteria for the base geospatial data themes and the custodians and has identified 10 base themes and their custodians: administrative boundaries, imagery, roads, social statistics, land use, land cover, cadastre, hydrology, geodesy and conservation. While specific geospatial digital data sets have been identified that are complete and available, much work still needs to be done to identify the feature types (classes) and attributes that should be included in the themes as base data. SASDI needs to cater for the peculiarities of the South African environment: legal, political, social, technical skills, technologies, etc.

The CSI has also developed its programme of work, established six sub-committees (for policy and legislation; data; standards; systems; education; and communications and marketing), drafted several policies, prepared a draft National Geoinformation Management Strategy, initiated negotiations for a site licence with the South African Bureau of Standards (SABS) for the relevant South African National Standards, conducted a study on the supply and demand of skills in geographical information science (GISc)[4] and the availability of suitable training material, arranged various workshops and other outreach events, and in collaboration with the South African Environmental Observatory Network (SAEON), set up a pilot metadata catalogue for testing the capturing and publishing of metadata by the custodians. The CSI is also contributing to the United Nations' initiative on Global Geospatial Information Management (UN-GGIM) and members of the CSI and its sub-committees are active in ISO/TC 211, *Geographic information/Geomatics*, the Open Geospatial Consortium, Inc (OGC), and the Commission on Geoinformation Infrastructures and Standards of the International Cartographic Association (ICA).

Architecture of an SDI

A key issue with the architecture of any SDI is whether it should have one centralized repository of all the geospatial data, or if the geospatial data should remain within the decentralized repositories of the relevant data custodians, with the SDI linking them together. Clearly, those who want to control the data, for whatever reason, prefer a centralized architecture. However, such an SDI with a centralised physical repository of all its data (even if it is in the cloud) will fail, simply because of the costs and complexities involved, never mind the inevitable lack of domain expertise at the central site. It will also be a single point of failure. In any case, the cost, reliability and bandwidth of the networks in South Africa make a central repository infeasible.

An SDI does not replace existing systems, but provides a one-stop data discovery and evaluation portal (though metadata) that reaches across organisational boundaries, and then routes queries for data sets directly to the relevant data custodians for retrieval. It is a *virtual* repository of all the data. A national SDI encourages the development of geoportals by national and provincial departments, local authorities and other agencies.

An SDI can provide web services for aggregating data sets and presenting them to the user as a seamless, integrated data set. It can also provide services for analysis, projection transformation, portrayal, validation, etc. An SDI obtains its metadata by harvesting automatically the metadata from its data custodians, or when a data custodian submits the metadata to the SDI.

An SDI should also provide easy access to the standards, taxonomies, terminology, ontologies and other things needed to exploit the data, such as through registries. An SDI could also contain a repository of orphaned, historical geospatial data sets (for which there is no longer a custodian). However, an SDI cannot contain a physical repository of current data sets. Each data set needs to be housed by its custodian, which will have the domain expertise and line budget to

[4] Unsurprisingly, there is a general lack of knowledge of SDIs in South Africa, with the result that in general, the CSI members lack expertise on SDIs.

maintain and update the data set and field queries about it, etc. A good SDI should also provide access through mirror sites around the country.

Commentary on current issues

1) There is much competition already to official or commercial SDIs, such as from virtual globes, open data repositories and the SDIs of other countries and regional organisations. Hence, any SDI has to offer a valid value-proposition to justify its existence. The SDI must also cooperate with other geoportals, to reduce costs and improve data quality and availability.

2) Studies in various countries have shown that volunteered geographical information (VGI) can be as good as, if not better than, official data sets. In particular, VGI data sets are often updated more quickly than official data sets, which adhere to an update cycle and have to provide national coverage, rather than just coverage of the "interesting" areas. Hence, VGI can supplement the official data sets to make any SDI more responsive and even more successful. An obvious use of VGI is for change detection [Guélat 2009].

3) The demand for conformance testing for, and validation of, data sets is increasing, particularly for dynamic applications such as in-vehicle navigation. Unfortunately, these are very expensive to do and hence would need extensive funding if they are to be a part of the offerings of any SDI.

4) An SDI that charges anything for online access to its geospatial data sets will fail. The harsh reality is that users expect to obtain data on the Internet for free (particularly as they have already paid for the data collection and maintenance through their taxes), and they will merely bypass any SDI that charges them for data. Countries such as Denmark have found that making geospatial data freely available generates far more in tax revenues than it costs to capture, maintain and disseminate the data [DECA 2010].

5) Implementing metadata is often not treated seriously, apparently because it is perceived to be tedious and expensive to capture and because there is no perceived value to the capturer – though obviously, anyone capturing data needs to document the data for their own use.

Recommendations

1) No SDI is an end in itself: It should support effectively and efficiently, evidence-based planning, participatory democracy, service delivery and sustainable development in South Africa to improve the quality of life of all.

2) An SDI needs a data acquisition policy, taking into account existing geospatial data in identifying the core data sets; harmonizing and planning data capture; setting data maintenance cycles; quality assurance of data; data interoperability and data integration.

3) An SDI should be implemented through pilot projects, to test the technologies. Thereafter, it can be scaled up gradually. Invariably, everywhere in the world, the "big bang" approach results in failure, leading to resistance to the system.

4) Because there are many unknowns related to implementing an SDI, particularly in a country such as South Africa, SDI research needs to be funded, both to address issues that arise and to develop advanced SDI skills in South Africa. Uncertainties that need to be researched for any SDI in South Africa include:

a. How to ensure it actually will support effectively and efficiently, evidence-based planning, participatory democracy, service delivery and sustainable development in South Africa, to improve the quality of life for all?

b. What the characteristics of the South African environment are that will make the SDI either easier or more difficult to establish and sustain than SDIs in other parts of the world?

c. How to ensure access to all, including people with disabilities?

5) Similarly, funding is needed for preparing suitable education and training materials, getting formal accreditation and presenting courses, etc. Training is required by various stakeholders in any SDI in South Africa, such as those tasked with implementing the SDI, domain experts, data custodians, data base administrators, developers of SDIs, the end users; etc.

6) Each data custodian from all three tiers of government and from outside government will have to have a substantial line budget for their component of any SDI.

7) The standards needed for any SDI to succeed need to be identified, eg: data quality, metadata, ontologies, data models, unique identifiers, software interfaces, web services, network services, encoding and portrayal.

Sources consulted

Clinton WJ (1994) *Coordinating geographic data acquisition and access: the national spatial data infrastructure*. Executive Order, 12906:pp 17 671–17 674.

Cooper, Antony K & Peled, Ammatzia (Aug 2001) Incremental updating and versioning. In: 20th International Cartographic Conference, vol 4, Beijing, China, pp 2804–2809. URL http://cartography.tuwien.ac.at/ica/documents/ICC_proceedings/ICC2001/icc2001/file/f19007.doc.

DECA (7 July 2010) The value of Danish address data: Social benefits from the 2002 agreement on procuring address data etc. free of charge. Danish Enterprise and Construction Agency. http://www.adresse-info.dk/Portals/2/Benefit/Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf

Gauteng City-Region Observatory (2013) Home page. http://www.gcro.ac.za/

Guélat, Jean-Christophe (20–21 Aug 2009) Integration of user generated content into national databases — revision workflow at swisstopo. In: 1st EuroSDR Workshop on Crowd Sourcing for Updating National Databases, Wabern, Switzerland. http://www.eurosdr.net/workshops/crowdsourcing_2009/

Hjelmager J, Moellering H, Delgado T, Cooper AK, Rajabifard A, Rapant P, Danko DM, Huet M, Laurent D, Aalders HJGL, Iwaniak A, Abad P, Düren U & Martynenko A (2008) *An initial formal model for spatial data infrastructures*. International Journal of Geographical Information Science, 22(11):pp 1295–1309.

Nebert DD (2004) *Developing spatial data infrastructures: The SDI Cookbook*. Global Spatial Data Infrastructure Association (GSDI). URL http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf

Schwabe, Craig A & Govender, Sives (May 2012) Stakeholder survey on defining the criteria and identifying core geospatial datasets and data custodians in South Africa. AfricaScope and EIS-Africa. Technical report prepared for the Development Bank of Southern Africa and the Committee for Spatial Information.

South Africa (2000) Promotion of Access to Information Act (Act No 2 of 2000). G 20852, 3 February 2000, Government Printer. URL http://www.info.gov.za/view/DownloadFileAction?id=68186.

South Africa (2003) Spatial Data Infrastructure Act (Act No 54 of 2003). G 25973, 4 February 2004, Government Printer. URL http://www.info.gov.za/view/DownloadFileAction?id=68032.

For more information on this policy note, please contact*

Antony Cooper
CSIR Built Environment
acooper@csir.co.za
Tel: 012 841 4121

Suggested citation: Cooper AK (2013) *Spatial data infrastructures (SDIs) and observatories for spatial analysis*. Policy Note 2, Spatial and Temporal Evidence for Planning in South Africa (stepSA), Pretoria, South Africa.