The National Data Grid: a Development Model for Grid Cell Data Infrastructure

Tai CHAN, Stephen FARRELL and Glenn FRANKISH, Australia

Key words: nested, grid, raster, grid-cell, national, data, infrastructure

SUMMARY

The National Data Grid (NDG) is a spatial data infrastructure (SDI) of grid cell (also called raster) data that is designed for both data custodians and users across Australia. The development of NDG is a process of knowledge generation, application and diffusion within the Australian Innovation System.

Cooperative Research Centre for Spatial Information (CRCSI), as an element in the Australian Innovation System, is adopted by the NDG project as a mechanism for collaboration among its key investors/participants. The benefits of the CRCSI as a collaboration mechanism is described in the context of Rogers’ organisation innovation theory and an approach of NDG development is identified.

The importance of promoting reinvention through the demonstrator program of the CRCSI and active engagement with the participants in maintaining the support for NDG, and thus, promoting diffusion, is highlighted. This is followed by a review of the achievements of the project to date and a pointer to the strategy of future development.
1. INTRODUCTION

The National Data Grid (NDG) is a spatial data infrastructure (SDI) of grid cell (also called raster) data that is designed for two main user groups across Australia. NDG aims to support the two groups and reduce duplicated investments by making it easy for:

- the data custodians to publish and share their data in a grid cell specification that is in an acceptable (authoritative) form suitable for general consumption
- the users of grid cell data to discover, access, integrate and use the data in their daily activities, eg., modelling, analysis, reporting and decision making in general (FGDC 2000).

The success of a SDI such as NDG should see hundreds, if not thousands of attributes, published and regularly updated on infrastructures across Australia, which is efficiently and effectively managed by one or more sponsors and accessed by thousands of users across Australia daily when undertaking their routine activities. It is an innovation that requires multi-million dollars investments over time by governments and the private sector, which would be impossible to come by without broad adoption by both the data custodians and users, and their respective organisations. The successful development and adoption of an innovative SDI such as NDG will vary in countries that will have different innovation systems shaped by their socio-political, economic and environmental contexts. The paper documents the model of development and diffusion of the NDG in Australia.

2. THE AUSTRALIAN INNOVATION SYSTEM

According to a review of the national innovation system (Australian Government 2008), an innovation system is made up of a range of components that include entrepreneurial firms and innovative work place, people and social networks, research capability and platforms, tax incentives, innovation in government and the innovation governance, policies and means to support all the above. A system also has three functions: knowledge generation, knowledge application and knowledge diffusion and benefits are sustained through the above components taking part in these functions.

Among the myriad of mechanisms available in the Australian innovation system, the Cooperation Research Centre Program is an important component. A recent review of the Program (Minister IISR 2008) recommended support for pre-competitive or pre-applicative research ventures between end-users and researchers which tackle a clearly-articulated, major challenge for the end users addressing identified risk gaps. The review also emphasised the importance of collaboration between end users and researchers and recommended improved...
support to collaboration through a number of existing mechanism, such as the Australian Research Council grant system, particularly with small and medium entreprises.

While the CRC program provides a effective means of supporting collaborative research to generate and apply the knowledge, the ultimate challenge in an innovation system is to promote knowledge diffusion, i.e., uptake and utilisation by the wider community. The CRC program facilitates that through education and demonstration programs. Education is to create capacity to assess, appreciate and use the innovation. Demonstration is to raise awareness and allow end users to see for themselves the benefits.

In the national innovation review it is acknowledged that enabling technologies such as Information and Communications Technologies (ICT) have widespread applications in many fields. SDI such as NDG is a typical example. Its use and management is deemed to involve significant complexity and uncertainty because it does not “fit into existing industry categories” and the end users cannot predict the outcome of their adoption based on past experiences. Therefore while a new bushfire behaviour model, once proven to be accurate and practical, will be adopted by fire managers, it is not the case for innovations such as NDG. It is an infrastructure that is designed to support a range of functions in various application areas. Positive experience in one area may not be seen as workable elsewhere.

The review further suggests that governments be “informed about enabling technologies and aware of potential issues and problems to develop appropriate policies and regulation” to cultivate the community’s confidence in a new technology and its subsequent acceptance and diffusion into the broader economy and society. It also highlights that “the community also needs to have access to balanced and objective information from trusted sources” to allow their choices to be informed. To address these challenges, strategies derived from organisational innovation theory can help (Chan et al 2008). The rest of the paper documents the strategy adopted and the outcomes.

3. INNOVATION MANAGEMENT: THE NATIONAL DATA GRID

The National Data Grid refers to the Australian Cooperative Research Centre for Spatial information (CRCSI) Demonstrator Project, which started in February 2007 as the Platform for Environmental Modelling Support (PEMS) Demonstrator Project (Chan et al 2007). The success of the PEMS demonstrator resulted in additional investments by the participants of the project and the CRCSI in September 2008. In this paper, the term NDG refers to both the current and previous demonstrators, i.e., PEMS, and the associated technical platform, unless specified otherwise.

The NDG was originally conceived very simply as an infrastructure with spatial information key as attributes against unique grid cells of consistent resolutions. The resulting data would be stored in simple relational databases to support sieve mapping type of queries to support rapid site selection analyses in the planning of major capital projects, such as building a prison or a toxic waste dump. With a critical mass of attributes in the grid cell database, it would also provide modellers with easy access to current grid cell data to feed their
sophisticated models. These models were developed to generate socio-economical and environmental predictions to inform decision making. The primary rational for developing the NDG was the reduction of duplication of valuable project budgets in data acquisition that often took up 75% of the budget of a GIS project (FGDC 2000). A small project was commissioned to create a proof of concept to communicate the ideas to the potential partners of the NDG.

Each of the above design elements of the NDG is not new and might have been applied individually to many organisations across Australia. The combination of all the elements in the same infrastructure is innovative requiring investment in new knowledge and its application (see Australian Innovation System above). For the NDG to be adopted as the mainstream component of the Australian SDI, adoption by many key organisations across Australia is a prerequisite. Promoting this adoption, or in the terms of the Australian Innovation System, knowledge diffusion of NDG, is a complex and challenging process.

As part of the strategy to address the above challenges, the concept, originated from Spatial Information Infrastructure, was presented to many business units in a number of organisations, including Geoscience Australia, the Australian Department of Agriculture, Fisheries and Forestry, the Victorian Department of Sustainability and Environment (DSE), Department of Primary Industries (DPI), Department of Justice (DoJ) and other agencies. Some private companies who were seen as potential developer-partners were also approached. Most business units approached were interested in the concept but did not see immediate value to participate in the project. They either did not understand the full potential of the project, felt that the timing was not right or simply too busy to consider new options. In the end, five government business units signed up. They were Geoscience Australia, the Economics Branch and Spatial Information Infrastructure (SII) of DSE, the Office of Emergency Services Commissioner of DoJ and the Future Farming Systems Research of DPI. A private sector company, Spatial Vision participated as a developer-partner.

In Australia and particularly, the state of Victoria, there are many mechanisms in the innovation system by which the NDG could be funded. They include the Australian Research Council Linkdage Grant, an appropriate Cooperative Reserach Centre (CRC), the various grants under the Victorian Innovation Statement (2008), or just simply joint investment by two or more like-mind partners. In the end, CRC was chosen as the mechanism. This was prompted by the facts that Geoscience Australia, the Victorian government and Spatial Vision were all members of the Cooperative Research Centre for Spatial Information (CRCSI) and that the latter had a Demonstrator Program and a Commercialisation Program to promote knowledge application and knowledge diffusion respectively.

Based on Rogers’ (1995) diffusion paradigm, Chan et al (2008) describe a model based on the Rogers’ ‘Organisational Innovation Process’ theory, which explains the roles of the CRCSI Demonstrator Program and Commercialisation Program in developing PEMS, the predecessor of NDG, and promoting its diffusion. They also highlight the importance of good project management in keeping existing participants engaged through building consensus on the technical capabilities of the infrastructure and ensuring the capabilities provide options to
improve existing business processes. They also point out the changing identify of an innovation and the importance of reinvention in promoting diffusion of PEMS.

4. MANAGING REINVENTION AND THE CHANGING IDENTIFY OF NDG

After articulating the initial concept of NDG to the stakeholders and deciding to use the Demonstrator Program of the CRCSI as the collaboration framework to develop NDG, the partners jointly developed the project proposal for the CRCSI. This marked the first evolution of the identity in the translation of NDG from the original proof of concept into a prototype designed by and for the participants. This is a major milestone in the evolving identity of NDG (Chan and Williamson 1998). The end users of the prototype were identified as:

- Business staff, managers and other professionals who are GIS novices
- GIS specialists and application developers
- Professional modellers
- Data custodians

A preliminary technological design was included with the demonstrator project focusing on building the following modules for grid cell data management:

- Data development and updates
- Central data store
- Data definition and extraction
- Data visualisation and export

4.1 The First NDG Demonstrator

After securing $260,000 of CRCSI funding, the project started in February 2007 and proceeded to clearly define the use cases, their feasibility and the detailed technical and functional specifications of the NDG. On its completion in June 2008, the demonstrator project developed a prototype infrastructure making up of the four modules listed above built on the core technologies of ESRI ArcGIS Server 9.2 and MS SQL Server 2005 as illustrated in Figure 1.

The use cases nominated in the project proposal to test the value of the proposed NDG technical infrastructure were as follows.

- Data support for the Market-Based Solutions Instrument Project, a project in which users of a set of models called EnSym (DSE 2009) would be able to download the data they needed to assess the potential environmental value of certain land management activities
- Support strategic wildfire planning and analysis – to manage a set of 170+ Victoria-wide asset value attributes in grid cell format and to allow local fire planners to apply weightings to them to identify priority areas in local fire plans
- Support regional land use modelling and analysis – to experiment with processes to develop current jurisdictional land use dataset and to support the use of land use models in policy development
- Creation of a Land Health Index
- Support land use and ground water modelling and triple bottom line indicators analysis
- National mapping of invasive species
- Development of national land cover maps based of FAO classification
- National framework for classifying land use and other attributes for Environmental Account reporting.

Figure 1: NDG Technical Architecture: Publication Data Store and Web Interface.

In the course of more indepth discussions to better define the use cases, the relevant stakeholders identified additional hurdles that they could not overcome to bring about the use cases. The hurdles included lack of time and staff resources, timing of the NDG did not fit in with current projects, lack of readiness conceptually in applying the NDG technology to the issues in question. As a result, only the first three use cases were developed and tested.
At the same time, other participants were finding new use cases. Spatial Information Infrastructure (SII) invested in two additional use cases. One was to develop an inter-operable (web services) linkage between its existing data ordering application, Spatial Datamart, to demonstrate the interoperable design of NDG. The other was to demonstrate the concept of applying the generic query function (sieve mapping) of NDG in rapid site selection analysis. Geoscience Australia also invested in a use case to manage time series data of Landsat derived normalised difference vegetation index, also called the “greenness index” which was then used to generate charts and maps within NDG to demonstrate the value of grid cell data management in supporting crop productivity prediction by the Australian Bureau of Agricultural and Resource Economics (ABARE).

Of the six use cases actually implemented, only the first two and the one involving land suitability analysis were successful in demonstrating the potential value of NDG. The technology failed to meet the requirements of the third use case, regional land use modelling and analysis. This was in spite of the effort made by both the sponsoring participant, DPI, and the developer, Spatial Vision to overcome the constraints imposed by the chosen technology to increase the volume of data that could be extracted in each data order. The use case involving the development of inter-operable linkage came out with inconclusive outcome as the Spatial Datamart application adopted an out of date open standard specification. The last use case involving crop productivity prediction failed to arouse the interest of ABARE.

The above experience suggests that even with the true believers, the participants of the NDG project, who actually put in the time and effort to shape the development of the prototype infrastructure, there was no guarantee that the outcome would meet their expectations. The demonstrator allowed the participant to reinvent the solutions offered by the NDG. The concept of reinvention was developed to Rogers (1995) describing a process in which an innovation is modified to fit the organisation in question. The more successful the reinvention is, the stronger will be the commitment to adopt the innovation in the organisation. In the case of NDG, successful reinvention by the participants were limited by the technologies and specifications chosen for NDG, the compatibility or maturity of the external applications or business processes supported, and the level of the end-users’ understanding (or insight) of the infrastructure and its implications.

The last limitation is highlighted by Chan et al. (2008) when they emphasise the importance of the adopter of innovation (NDG in this case) being able to experiment with and tweak the infrastructure to meet their precise requirements. This is also brought out in the Cutler Review (Australian Government 2008), which observes that management and use of enabling technologies such as Information and Communications Technologies, (ie., the grid cell data infrastructure of NDG) “involves an uncommon degree of complexity and uncertainty”. The end-users need “balanced and objective information from trusted sources so that they can make informed choices”. This is best achieved through regulations by governments (or in the case of NDG, the Project Control Group that governs its operation), personal experience or recommendation by trusted partners or collaborators. In the case of ABARE, their involvement was mediated through Geoscience Australia, making it hard for them to realise
the true value of NDG. At the time of the project, the lack of published statistics or regulations that govern the proper use of the infrastructure, might have increased the perceived risk of adopting NDG in their business process despite the success of the use case.

Based on the technical platform, the capabilities implemented were as follows.

1. a national nested grid system making up of two demonstration sub-systems: one based on national latitude-longitude coordinate and the other Victorian Vicgrid map projection; it was used for organising spatial information across Australia over a wide range of scales - local to national
2. a publication data store with the associated user interface and data model for functions that include visualisation: map, charts and reports; query; modelling by applying simple weightings to attributes; management of grid cell data by theme and time series; data search, selection, ordering, extraction and delivery
3. preliminary research for a separate backend data store designed for scalability and management of very large volume of data. The store was meant to support the publication data store that was designed for efficient querying and visualisation
4. test data that was made up of selected national climate datasets (6 minutes resolution), jurisdictional (Victorian) statewide data for asset values (1 km) and land use and catchment assessment data for selected catchment management authorities of Malle and Corangamite (100 and 20 m).

From a functionality viewpoint and with a focus on the publication data store, project participants identified the key benefits of the first NDG Demonstrator application to be the provision of:

- Single point for data discovery and interrogation
- Significant functionality and
- Data bundling capabilities, especially in relation to data extraction.

The key concerns or issues raised by participants in relation to the performance of the first Demonstrator were:

- Extraction and spatial query/display performance issues relating to fine resolution grid data
- Web connectivity and performance issues for selected project participants
- Screen navigation, completeness and terminology and
- The capacity to handle very large datasets, i.e., the design and implementation of a high capacity and scalable backend data store
- The capacity to upload data easily and cheaply into the grid cell data environment.

Some participants were encouraged by the success of the first demonstrator and the lessons learnt. They also recognised that the first demonstrator was meant to be a prototype and not suited to supporting mission critical business process. The Victorian Department of Sustainability and Environment and Department of Justice decided to invest in a 3-year pilot
to evaluate the capacity of the improved infrastructure in supporting actual business processes. The participants as a whole decided to take the project to the next step and seek the support of the CRCSI to co-fund the infrastructure so that it could be kept as a demonstrating platform for other interested parties, particularly government departments.

4.2 The Second Demonstrator

Leveraging off the investment by the participants, a proposal was put to the CRCSI to fund an additional demonstrator project with a view of addressing the key issues identified by the first demonstrator and to make the infrastructure available for testing/experimentation by other participants, existing or new. At the same time, another CRCSI participant, Victorian Partnership for Advanced Computing (VPAC) from the private sector, expressed interest to join the project. The participants recognised the value of VPAC’s expertise in high performance computing and in managing very large databases for industrial applications in, say, automobile design, in further developing NDG to meet the needs of the wider community. As a result, based on inputs from the participants, VPAC and the CRCSI, a $880,000 demonstrator project was funded to achieve the following objectives:

1. to upgrade the existing NDG infrastructure to production standard to support business processes nominated by anchor tenants and to support a more thorough evaluation not possible in the current demonstrator
2. to support ongoing demonstration of the value of NDG to potential national and jurisdictional agencies and the public
3. to investigate innovative means to address scalability issues.

In particular, the third objective was about developing a strategy and selected innovative front end and back end capabilities to address scalability issues faced by the NDG. The project proposal helped to document the updated collective identity of the NDG as perceived by the participants. It also represented a major step in the reinvention of NDG to suit the emerging needs.

After formalising the governance arrangement, the participants tried to flesh out the detailed identity of NDG and decided that it would be made up of two separate components. They are:

- improved infrastructure of the current NDG to support a 3-year pilot for the key investors, i.e., DSE and DoJ, in a production setting, to be undertaken by Spatial Vision
- a prototype of a scalable high capacity backend data storage environment to be undertaken by VPAC, which would support the current NDG in due course.

After extensive research, testing and consultation, VPAC came up with a design for the back end data store which was approved by the Project Control Group and named Raster Storage Archive. The broad design is summarised as follows and illustrated in Figure 3.
Data Storage:
- Image data formats to optimise storage capacity
- Lossless compression to maintain data integrity
- Tiled data storage to optimise retrieval performance
- Australia-wide multi-resolution index/indices for unique identification, re-sampling of grid cells and for future integration with the existing NDG infrastructure

Hardware/software:
- Open source software/tools, in particular the core raster data management tool of Geospatial Data Abstraction Library (GDAL) to reduce licensing overheads
- System deployed as server instances in a corporate virtual server environment for ease of deployment and scaling up
- Load balancing configuration to support scaling up
- Federation of RSA nodes through exchange of index information to maximise custodial control of data at source.

Figure 3: Proposed NDG Raster Storage Archive Scalable Configuration.

As the consultation progressed, the participants began to understand what the Raster Storage Archive (RSA) could do. They saw this reinvention of the NDG offer a better match to their data management needs that involved storing, processing and distributing national and victoria-wide datasets at comparatively high resolutions. Geoscience Australia decided to test the management of five attributes of the national radiometric and gravimetric dataset at a scale of 100m while DSE the group of eleven input datasets for the Victorian EnSym suite of models at a scale of 20m and had data ordered delivered in the EnSym specific “Matlab” format. DSE also invested in building a web service link to allow orders for data upload and download to be placed directly within the EnSym user interface, further automating the transaction between EnSym and the NDG RSA. The new opportunities identified by the two organisations were deemed out-of-scope in the specification that guided improvement of the
existing NDG. By this time, the latter was referred to as the Publication Data Store, the term used in the original design document of the first demonstrator.

At the same time, both Geoscience Australia and DPI identified opportunities to make use of the Publication Data Store (PDS) to meet specific data management needs. In conjunction with the Bureau of Rural Sciences of the Australian Department of Agriculture, Fisheries and Forestry (DAFF), Geoscience Australia invested in a study to manage a group of attributes needed by a national modelling package called the Multi-Criteria Analysis Shell for Spatial Decision Support (MCAS-S) (Bureau of Rural Sciences 2009), to support decision making by the Queensland rural land managers in their bid to better manage the Great Barrier Reef. The project is called Reef Rescue. The developer of PDS worked closely with the Reef Rescue project team to identify an optimal cell resolution to allow the project to make full use of the functions of PDS, assisting them in the reinvention process to adopt the innovation.

DPI, on the other hand, identified a need to manage the dozens of census and agricultural economic attributes that had to be used in an interactive community engagement process to determine an index for Adaptive Capacity to Climate Change for the Southwestern part of Victoria. This process saw value in the weighting function offered by the PDS to allow the community representatives to interact with the system in real time to generate the index. DPI was keen to evaluate the benefits of PDS in this process.

5. THE WAY FORWARD

With the reinvention process taking place in the background, the participants appreciate the potential of the two components of the NDG better. The Project Control Group has instructed that the identity and value of the NDG be properly articulated as a holistic entity. As a result the project got better in articulating the value of the NDG to potential adopters of the innovative infrastructure. There were two recent examples.

- **Urban Digital Elevation Model (UDEM).** After comparing the roles of the UDEM project (Fraser and Wheeler 2008) with that of the NDG, representatives of the Steering Committee of the UDEM project agreed to jointly research into means of collaborating with the NDG project to diversity the number of public channels to access the upcoming authoritative national DEM products.

- **One-stop-shop for Fire Models Integration.** Being a participant of the NDG Demonstrator project, representative of the Office of Emergency Services Commission (OESC) on the Project Control Group had an in-principle agreement with the project to research into ways of supporting various fire and other emergency risk models in Victoria. These include providing input data support to the models and as a repository to share and visualise fire risk model outputs efficiently.

When realised, the above collaborative research activities are expected to generate new insight into the grid cell data needs of the wider community through the associated reinvention of the NDG. The resulting support of the wider community of users would
strengthen the case for more sustainable fundings from various government sources, including funding from existing innovation programs nationally and in Victoria.

In the course of undertaking the two demonstrator projects, stakeholders, both existing and potential, have identified the needs for appropriate regulations or governance processes to ensure data published are up-to-date and to the specifications of the data custodians. This is to ensure that the data published are regarded as authoritative and “fit-for-purpose”, and will require the codification and implementation of the best practices for information management. This demand for appropriate regulations is in line with the measures recommended to promote diffusion of enabling technologies in the Cutler review (Australian Government 2008). It is expected that this will have to be dealt with in reinvention activities in the future to ensure that the NDG stays relevant to the adopters and end users.

With the success described above, the logical strategy to take the NDG infrastructure forward for wide adoption across Australia involve the following steps:

- Continue to work with existing participants to ensure that their existing and emerging grid cell data management needs are well articulated and met
- Actively engage with new/potential participants, including data custodians, users and system developers, through the extended network of contacts of the existing participants to extent capabilities of the infrastructure to stay relevant
- Continue to maintain a demonstration component of the project to allow reinvention and diffusion to take place both for existing and new participants
- Robust and transparent project management to manage the trust and expectation of the participants
- Continue to make use of existing mechanisms to facilitate collaboration and diffusion within the Australia Innovation System, especially the CRCSI-2, which has just secured $32 M of funding from the Australian Commonwealth for another eight years starting 1 January 2010.

6. CONCLUSIONS

The National Data Grid is an important component of the Australian Spatial Data Infrastructure, helping to meet the grid cell (or raster) data needs of the nation, particularly in modelling. While the technologies are not new, the way they are put together to provide relevant services to the range of users concerned is innovative.

In the innovation system in Australia, the Cooperative Research Centre for Spatial Information was chosen as the collaborative mechanism to generate and apply the knowledge of the design and management of the National Data Grid infrastructure. This is done through two separate demonstrator projects part funded by the CRCSI. However it still takes good project management and stakeholder engagement to ensure diffusion of the infrastructure across the potential adopters in Australia, especially among the participants of the projects. An important approach is to promote “reinvention” of the infrastructure to ensure that it fits the organisations that are going to adopt it.
REFERENCES

BIOGRAPHICAL NOTES

Tai On (T.O.) Chan is currently Principal Spatial Policy Officer, responsible for the spatial policy and research matters in the Department of Sustainability and Environment, State Government of Victoria, Australia. He had a degree in Biology and in Forestry. From 1983-94, he worked for the Hong Kong Government as a Forestry Officer with responsibilities of planning, conservation, management and training inside country parks. He obtained his PhD at the University of Melbourne in 1999. His professional and research interests are in planning, management and implementation of spatial information systems and spatial data infrastructures.

CONTACTS

Dr Tai On (T.O.) Chan
Department of Sustainability and Environment
P.O. Box 500
East Melbourne, Victoria
Australia
Tel. +61 3 8638 2682
Fax + 61 3 8636 2813
Email: tai.chan@dse.vic.gov.au
Web site: www.dse.vic.gov.au