

Platform for Environmental Modelling Support: Design and Values of a Demonstrator of Grid Cell Data Infrastructure

Tai, Chan¹; Stephen, Farrell²; Dmitar Butrovski³;

¹Cooperative Research Centre for Spatial Information, Department of Sustainability and Environment, Melbourne, AUSTRALIA

²Cooperative Research Centre for Spatial Information, Spatial Vision, Melbourne, AUSTRALIA

³Cooperative Research Centre for Spatial Information, Geoscience Australia, Canberra, AUSTRALIA

Abstract

Governments worldwide and in Australia are relying increasingly on scientists, including environmental modellers and GIS analysts, to provide the evidence needed to respond to climate change and other environmental challenges. The scientists, in turn, require extended data management support from governments, particularly that for grid cell data. The PEMS Demonstrator project, sponsored by the CRCSI and supported by ANZLIC and ASIBA, is a prototype for a grid cell data infrastructure of the Australian Spatial Data Infrastructure that can provide extended data management support to the modellers and analysts. The design, innovations and value of the PEMS Demonstrator are described in the context of a typical life cycle of data utilisation. This paper then concludes by highlighting the strategic and tactical benefits that PEMS can offer.

Introduction

The Stern (2006) Review on the Economics of Climate change was published in October 2006. The last of a series of 4 reports by the Intergovernmental Panel on Climate Change was published in November 2007 (IPCC 2008). These and other events led the new Australian Government to ratify the Kyoto Protocol in December 2007. This confirms Australia's commitment to face up to the environmental challenges posed by climate change. Key responses include investing in research into climate change and developing adaptation strategies, particularly in terms of long term and emergency response to fire, drought and other extreme weather patterns such as frost, flood and intense storms. Underpinning these responses are the spatial information and knowledge of the environment needed to support decision making and policy development.

To support these key environmental responses, governments worldwide are developing spatial data infrastructures (SDIs) (Onsrud 1999). A good example is the INSPIRE (Annoni and Craglia 2005) initiative of the European Union. An SDI provides the standards, technical, institutional and policy means to facilitate access to spatial information (see Nebert 2004 for a detailed definition). In Australia, the effort is called the Australian Spatial Data Infrastructure (ASDI).

Grid cell data

A quick survey of the Australian Spatial Data Directory (ASDD 2008) of ASDI was conducted. It involved searching the ASDD using the key words: "ASCII" "vector", "point line polygon vector", "grid", "raster", "imagery" and "grid raster imagery", and compared the seven sets of figures with the total datasets reported in the quarterly report for Jul-Sep 2007 (ASDD 2008). The quick research indicates that there are significant data holdings in ASCII, vector, grid cell/raster, imagery. However as the terms are sometimes used interchangeably and in a number of metadata attributes, it is not easy to know the exact proportion of each data type/format held in the ASDD.

The value of grid cell data is well known to environmental (including agricultural) modellers and is summarised in an information sheet produced by the Cooperative Research Centre for Spatial Information (2007). A grid cell data structure is a matrix where the coordinate of a cell (also called pixel) can be calculated if the origin point is known and the size of the cells is known. Such a matrix-based structure lends itself to manipulation via two-dimensional arrays in computer encoding and as such allows for considerable efficiencies in the development of analytical (modelling) processes. Most grid cell based GIS software requires that each grid cell contain only a single discrete value, thus generating single attribute maps. This is in contrast to most conventional

vector data models that maintain data as multiple attribute maps, e.g. forest inventory polygons linked to a database table containing all attributes as columns. This basic distinction of grid cell data storage provides the efficient foundation for quantitative analysis techniques based on sophisticated mathematical modelling processes. This makes cellular (tessellated) data structures the paradigm of choice for the vast majority of environmental modelling processes, particularly those involving continuous data, e.g. elevation, weather data, ground water and vegetation distribution. On the other hand, it is noted that vector data systems are better suited to linear data analysis, e.g. shortest path, and high quality cartographical requirements.

Typical life cycle of data utilisation

To meet their grid cell data needs at the project level, environmental modellers and GIS analysts typically go through the following steps: design, search/acquire, process, integrate, analyse, report and publish (adapted from NDN 2004). The steps are briefly described below under the respective headings.

- *Design.* Based on their analytical needs, the modellers and analysts will identify the attributes, formats, datum and projections, times and locations of the data required.
- *Search/Access.* They will try to search both data published through the ASDD or other sources, and unpublished data through their contacts and make arrangements to access the datasets required. In case of unpublished data, the associated metadata may not be available, and if available, not up-to-date as the data may have been modified by successive users of the original data. In many cases, modellers and analysts are prepared to use any data that is available to get a result.
- *Process/Integrate.* Once they have the raw data, often they have to translate the data into a format suitable for their tools and models. This translation include activities such as datum, projection and scale conversion, format conversion in terms of both data type and software format, re-sampling of data if it is already grid cell based, and derivation of the attributes needed from the source data. They may have to derive a number of attributes for a specific geographical extent or one or more attributes for all the locations that suit a particular set of criteria.
- *Analyse and Report.* This is probably the simplest of all the steps as the modellers and analysts probably have better control over it. It involves running the data through the models or analytical processes developed to get an output which is then turned into a report. If the work is ongoing, then there is the effort to keep the attributes derived up-to-date, ie., going through the same process again on a regular basis.
- *Publish.* The final step of the data utilisation life cycle is to publish new data that are derived from their modelling and analyses so that they can be made available together with their source data. This allows other users to access both types of data to more innovative analyses that target specific environmental challenges. This is particularly important when dealing with climate change. With the scale of the challenge and the limited available, it is vital for governments to facilitate integrated research among scientists from different disciplines, ranging from climate change through hydrology, ecology, economy to agriculture, geography and planning, The strategic benefit here is to allow research outputs from one discipline to be used and built upon progressively by other disciplines and vice versa to fast track the development of integrated and more holistic solutions.

To certain modellers and analysts, the main frustration is encountered in the *Publish* phase. As custodians of the data outputs of their work, there is no easy way of publishing the data for reuse by other modellers and the community at large. This includes having the necessary storage capacity, means of discovery and access, generating the metadata and providing means of controlling access to sensitive data that is subject to interpretation.

At a jurisdictional level where data coordinators are charged with compiling comprehensive, consistent, correct and current jurisdictional datasets to support environmental decision making and policy development, the above project-based approach of data generation poses a different dilemma. While it is good to have a large number of modellers and analysts to generate the data layer needed across the jurisdiction, there are often variations in the way the project data are generated. Data coordinators are often faced with the challenge of trying to keep track of relevant

projects and combine project data that are different in various ways. Inevitably they have to go through the data use cycle similar to modellers at the project level to get their jurisdictional datasets, resulting in more duplication of resources.

Grid cell data infrastructure

The ASDD provides valuable services of search and discovery of published datasets to support environmental modellers and analysts. Sometimes they may extract easily the necessary attributes of the environment from the datasets identified in ASDD. Very often, these people have to hunt for other sources of data and undertake extensive data processing and conversion to get the attributes they want, which add to their overheads. Once they have done their analyses, they will be asked to provide/modify their outputs for other modellers and analysts. In short they need more support in the typical life cycle of utilisation of grid cell data, as described above, in order to allow them to contribute more efficiently and effectively to the governments' environmental agenda. They need a grid cell data infrastructure that offers services beyond the ASDD.

In this context the grand vision for the Australian grid cell data infrastructure is to facilitate access to grid cell data in Australia. It does so by developing the four key components of a SDI, ie., the data, standards, technology and governance (Nebert 2004). Data here refers to grid cell data. Standards are agreed technical specifications that facilitate integration and reuse of data. Technology is the combination of software, hardware and communication technologies (including the human resources) that provides the data services needed. Governance is the policy and institutional arrangements that lead to the development and upkeep of the other three components within and across organisations and jurisdictions.

PEMS – the CRCSI Demonstrator

To help meet the grid cell data needs of environmental modellers and analysts, the Victorian Government Spatial Information Infrastructure Branch, in concert with its counterparts in other sectors, formulated the idea of a grid cell data infrastructure of the ASDI almost 2 years ago. Through a rigorous process of stakeholder engagement, the idea was realized in the form of a Demonstrator Project called Platform for Environmental Modelling Support (PEMS). This project is sponsored by the Cooperative Research Centre for Spatial Information (CRCSI). The core participants include Geoscience Australia of the Commonwealth, the Department of Sustainability and Environment, Department of Primary Industries and Office of Emergency Services Commissioner of Victoria; and Spatial Vision of the private sector. ESRI Australia and Microsoft are the technology providers. The Demonstrator has the support of both the Spatial Information Council of Australia and New Zealand (ANZLIC), the Australian peak body for spatial information and Australian Spatial Information Business Association (ASIBA), the business association for spatial information.

In a Demonstrator, CRCSI participants use existing information and technologies in innovative ways to show the power of SI to existing or new users. This encourages adoption of SI and improves the awareness of the market about "spatial information" capabilities and applications. As such the primary role of PEMS is to show and evaluate the merits of the concept of a grid cell data infrastructure to key stakeholders. In this context the project has developed five use cases from selected business processes of participants. While in its scoping report the project has articulated the long term capabilities of PEMS as a production system, this Demonstrator will develop only a subset of these capabilities, automating some and mocking-up the remaining.

PEMS started in February 2007 and will take 18 months to complete. It has used the first six to eight months to set up its internal project governance and to ensure that all participants agree to the vision and scope of the project, functional specifications and the detailed implementation of the Demonstrator. At the time of writing, functional development is approaching completion and system testing is starting. An evaluation plan is in place to guide the evaluation of the merits of PEMS and will be implemented from March 2008 onwards.

The following sections highlight the compositions of the four key SDI components of PEMS, ie., data, standards, technology and governance.

Data

The data holding of PEMS is restricted to the datasets needed for the use cases. They include:

1. data in geographical (lat-long) grids

- Landsat Thematic Mapper (Raster) 1 sec (25 metres)
- MODIS Vegetation Index (NDVI) Time Series
- MODIS Vegetation Index (NDVI) Baseline Series
- Land use

2. data in Vicgrid projection grids (20/100m)

- Urban boundaries
- Land-use
- Land Use data set developed using the Australian Land Use and Management classification standard.
- Road network
- Slope
- Hillshade
- Exclusion (areas excluded from development)
- Forest Cover
- Irrigated Area
- CMA Regions
- LGA Mallee CMA
- DEM
- ASPECT
- CLIMPROX (location of weather stations)
- EVC (Ecological Vegetation Class)
- EVC1750 (EVC in the year 1750)
- RAINFALL
- TEMPERATURE
- GFS1M
- SOIL

3. data in Vicgrid projection grids (1 km)

- Economic Infrastructure Asset Class - Consequence of Loss
- Economic Production Asset Class - Consequence of Loss
- Environmental Biodiversity Asset Class - Consequence of Loss
- Social Infrastructure Asset Class - Consequence of Loss
- Social Cultural Asset Class - Consequence of Loss

Standards

In a presentation to ANZLIC during one of its meeting in 2006, ANZLIC gave its support for PEMS and indicated that it would consider adopting the relevant PEMS specifications as standards if they were proven successful (personal communication with Bruce Thompson, Victorian representative to ANZLIC). The key specifications developed for the PEMS Demonstrator which may be relevant here are the grid cell storage structure and a national nested grid. Both are developed in close consultation with core participants, especially Geoscience Australia, which, as the lead spatial organisation in the Australian Government, brought in views from other national organisations such as National Land and Water Resources Audit, Bureau of Rural Sciences, Department of Environment and Water Resources and CSIRO. The approach is to give the standards reality checks and the necessary credibility for acceptance by other stakeholders in Australia as the starting point for further development.

The data storage structure

Grid data has, by its nature, a very large volume, especially where large areas are covered by grids with small cell sizes (eg 20m grid for a state). Proprietary grid storage formats generally implement different compression mechanisms to minimise this data volume – such approaches however lessen the ability of applications to process this data quickly, especially when summing data across different datasets. Some specialist grid storage formats adopt a nested set of grids structure, where the one grid layer may have some areas represented at one grid size and other areas (where the data being represented varies more) at smaller grid sizes. Whilst this can also lead to significant saving in data volumes, the structure is more complex to manipulate using normal SQL queries as the relationship between grid cells in one location but across different layers must be spatially determined.

It is proposed that PEMS would support two separate grid structures, one for the storage of data and the other for the visualisation and querying of data. These two structures would be known as the Storage and Publication data stores. The Storage data store would be optimised for storage efficiency whilst the Publication data store would be optimised for drawing and querying speed. It is likely that some datasets may be loaded into the Storage data store but not into the Publication data store, hence leading to more efficient storage. It is also envisaged that a statewide dataset may be loaded into the Storage data store but only “published” for limited areas of the state and possibly only remain in the Publication data store for limited periods whilst a project is carried out. Figure 1 below represents this separation of data stores.

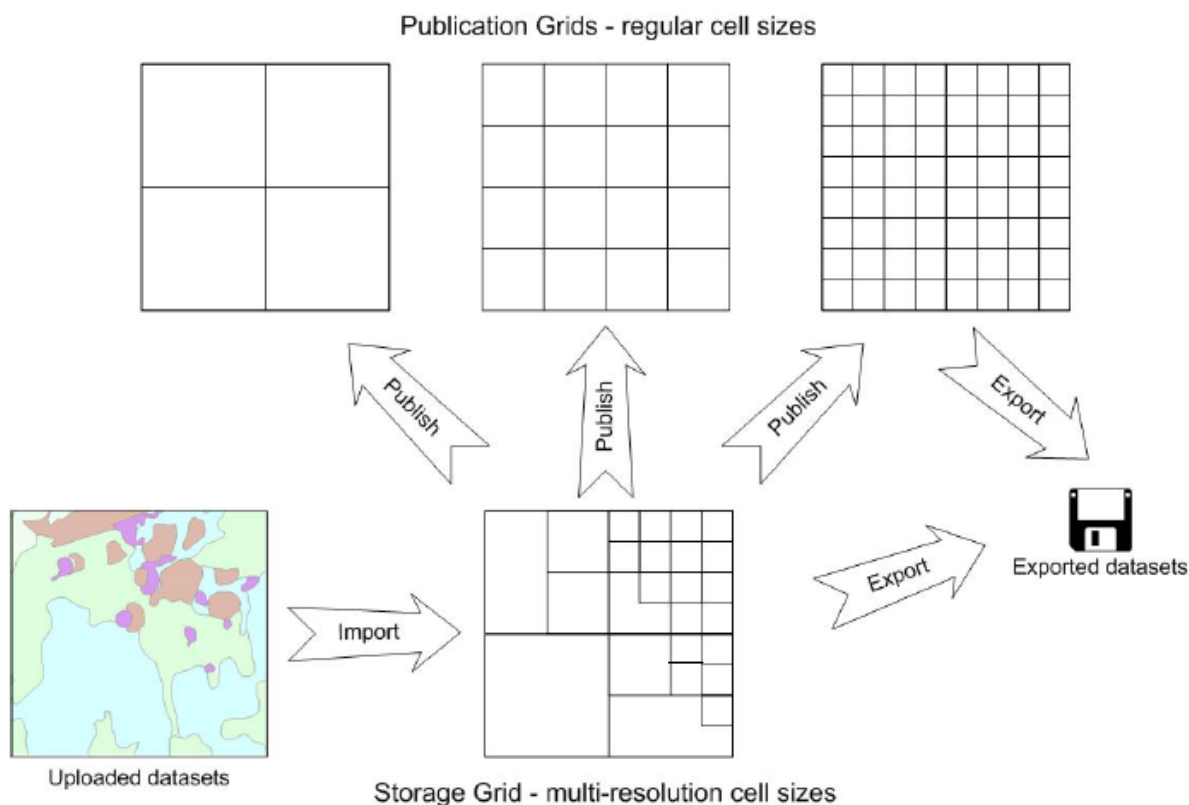


Figure 1. Representation of proposed separate grid structures, one for the storage of data and the other for the visualisation and querying of data.

Note that data may be exported from either a publication grid or the storage grid, depending on where a dataset exists and from which it is less work to generate the required export file. Note also that during the process of registering a dataset for importation, the data manager would be required to declare the levels, if any, at which the dataset can be published. It is possible that the publishing itself would be deferred, only occurring when a user actually requests access to that published dataset and grid level for viewing or querying. Such publication may also be able to be requested for a given area and period, with PEMS automatically cleaning up expired publications in order to conserve space. Certain commonly used datasets could be marked for permanent publication at various grid levels.

In practice the PEMS Demonstrator has implemented only the publication grid. The concept of the Storage data store has not been fully explored or defined. As part of the Demonstrator a separate process has been undertaken in relation to the storage grid to review its design and to assess performance, efficiencies and other practical implementation issues.

The Publication data store will comprise a nested set of grids from national to local scale. The data store will hold imported datasets in simple aspatial tables, with each attribute value identified on a location by location basis by a unique ID which relates it to a spatial location (point/grid cell). As well as the actual data values themselves, textual tables will be used to hold a range of metadata about datasets, users, organisations and so on to enable the operation of the application. The metadata are specific to PEMS, including items such as data of data upload, process of rasterisation used. They are meant to follow the ANZLIC metadata profile which, unfortunately, was not published at the time of development.

A set of static spatial datasets (point/polygon) will be used to enable the visualisation of datasets. To enrich the functions of PEMS, a small collection of spatial dataset is maintained to provide contextual data, such as topographic map or road networks, to allow better visualisation of the grid cell data.

In addition to being stored as spatial datasets, certain common spatial attributes such as local government areas, sub/catchments, mesh-blocks and postcodes can also be held in aspatial tables. This will allow “spatial” queries to be made through normal SQL queries even when spatial software is not operating or available.

The national nested grid – extent

The national nested grid has two aspects to it: the geographical extent and the detailed design of the grids in terms of cell sizes, datums and projections. Firstly it is the geographical extent that forms the Commonwealth of Australia. As a different datum is used for Antarctic Territory, it was agreed that data at the national level will be stored in two grids:

- 61-174°E, 8-61°S for mainland Australia and external territories excluding Antarctica; and
- 39-161°E, 60-90°S for Australian Antarctic Territory.

The extent of Australia’s jurisdiction, covering its Exclusive Economic Zone, is shown in Figure 2. The Demonstrator has implemented only the mainland grid.

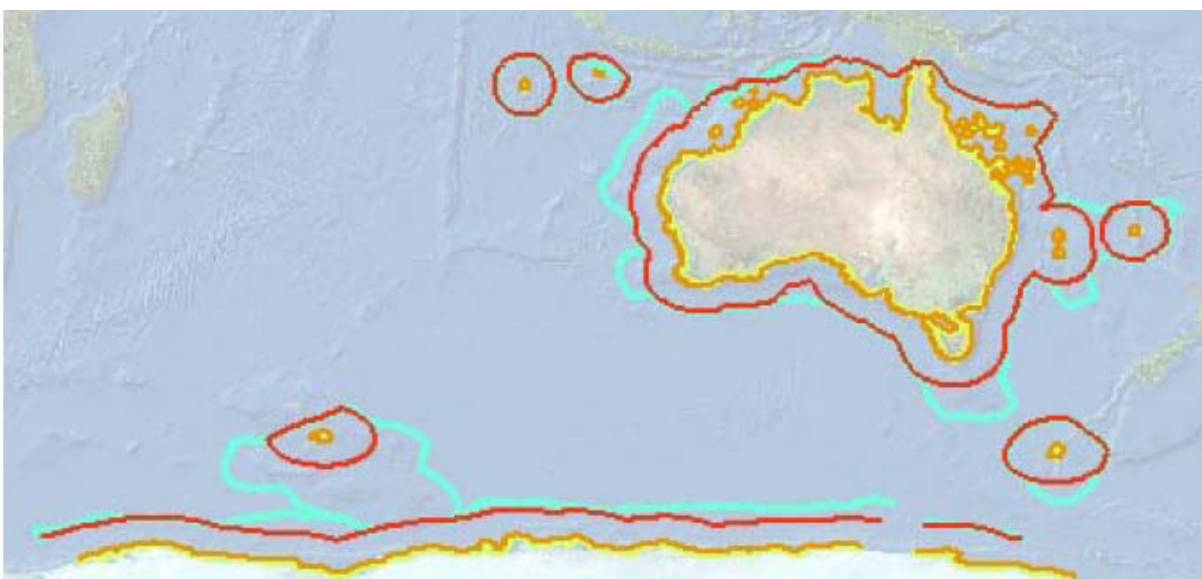


Figure 2. Australia’s Exclusive Economic Zone limit – red; Continental Shelf limit – cyan; Coastal Zone limit – brown; Territorial Sea - yellow

The national nested grid – projection and datum

The nested set of grids will comprise two grid systems. These include:

- Grid system 1 at the national level; and
- Grid system 2 at a state or sub-state/ jurisdictional level

Both grid systems will support multiple resolutions and are to take the form of a nested set of grids where the detailed grids are directly aligned and form a subset of the broader grid system. A separate nested set of grids will be implemented for each of the two grid systems identified. In grid system 2 a separate nested set of grids will be created for each Australian State or Territory.

Grid System 1 contains data stored in geographicals reference system, which supports analysis (and presentation) on a national basis. Its grid lattice aligned to the graticule formed by longitude and latitude lines. This approach will ensure information can be suitably processed, queried, and presented in a seamless manner. It will also ensure the national information (relevant to Australia's Exclusive Economic Zone (EEZ)) is suitable for integration with adjacent regions of the globe. The following national standards will be adopted:

- The Intergovernmental Committee on Surveying and Mapping (ICSM) resolved to adopt GDA94 as the national datum used for all non-military spatial data infrastructure. GDA94 is based on a global reference frame that allows transformation to any new realisation that may be produced in the future; and
- The Scientific Committee on Antarctic Research adopted the ANT2000 datum which is based on ITRF2000@2000 for its spatial data. As there is no land bridge or intersecting EEZ limits between the two areas, both datums can be used respectively.

The units of storage for geographicals grid information in Grid System 1 will be decimal degrees. The implications of distance and area measurements for grids stored in geographicals system are provided in Appendix 1. The national nested set of grids system implemented in the Demonstrator has a maximum resolution of 6 minutes (or approximately 10 Km) and a minimum of 4 seconds (or approximately 100m). Selected areas of this system may allow for 1 second (or approximately 25m) resolution. Typical applications of this grid system include the analysis and presentation of national climatic information, vegetation cover or fishery resources.

Grid System 2 supports the management of detailed spatial datasets, which involves the adoption of an agreed state jurisdictional projection and datum. In the case of the Demonstrator that only involves Victoria for the time being, this would be VicGrid and GDA datum. For other states/territories, the grid systems will be based on common practices adopted. A buffer strip of approximately 100km (or a suitable distance that equates to administrative or study area requirements) will be included within the nested set of grids generated for each jurisdiction. The Demonstrator will support a nested set of grids with a minimum resolution of 20m and a maximum grid size of 1km.

For state/sub-state jurisdictional datasets to be integrated and aggregated to an appropriate national dataset, it is crucial that a suitable translation framework is in place. In the Demonstrator both grid systems will include a grid of similar dimension. The 1km, 100m and 20m grid components of grid system 2 and the 36 second, 4 second and 1 second geographicals grid components respectively, in grid system 1 would serve this purpose. Presentation of information will not require data re-sampling, but rather an on-the-fly re-projection of the source grid data.

Technology

The technology PEMS is putting together is meant to support the spatial data utilisation life cycle. On the data supply side, PEMS supports the publication of existing or new data by data managers. Furthermore, as more scientists are engaged in modelling the environment, it is becoming more and more important to be able to allow modellers to published new data derived from running PEMS data through sophisticated external models to provide more sophisticated information about the environment. On the utilisation side, PEMS supports data exporters, querying/ reporting users, modelling users, mapping users and decision support system (DSS) users. To ensure all services run smoothly, PEMS must also support its administrators.

While the roles of data managers and administrators should be self-explanatory, the detailed roles of the users are summarised below.

- Data Exporters are data users who extract data from PEMS for use in external modelling and analysis tools.
- Querying/Reporting users are users who wish to make use of PEMS simple querying and reporting functions to meet specific business needs, routine or ad hoc.
- Modelling Users are users who wish to make use of PEMS mapping and simple modelling functions to derive new datasets for specific processes.
- Mapping Users are users who wish to undertake general viewing and simple querying of PEMS data, via the PEMS map interface.
- DSS Users are users who wish to use PEMS to get answers to well-defined (and possibly pre-defined) questions based on customised dataset queries.

To support the above five groups of people involved in the spatial data use cycle, the technical design of PEMS is based on the business function model illustrated in Figure 3.

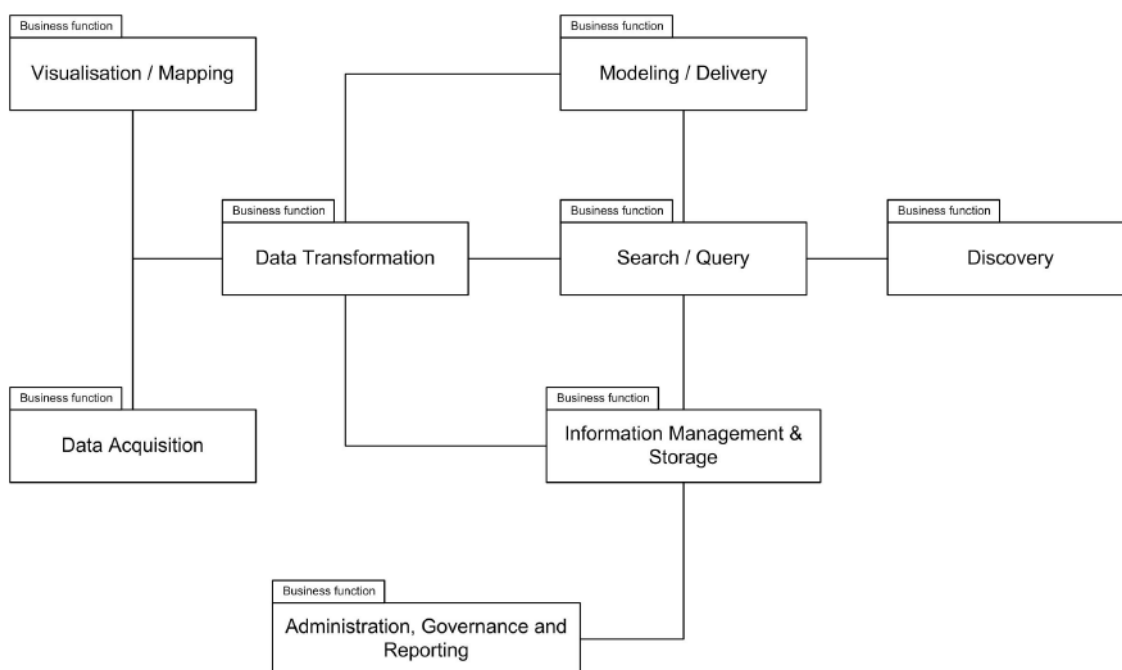


Figure 3. Business function model for PEMS Production

Of the eight business functions identified in Figure 3, theoretically, five can exist as an aspatial system. These functions are Info Management & Storage, Search/Query, Administration, Governance and Reporting and Discovery. However to meet the expectation of spatial information users nowadays, spatial technology has been incorporated into the technical design of PEMS to enrich the user experience. In any case, the functions of Data Acquisition, Data Transformation and Visualisation/Mapping need spatial technology to work properly. However in an extreme austere financial situation, an aspatial system is still possible.

The high level technical architecture adopted for the PEMS Demonstrator is illustrated in Figure 4. For the purpose of this paper, the key technologies chosen are as follows.

- Database: Microsoft SQL Server
- Spatial software: ESRI ArcSDE and ArcGIS Server
- Access software: common browsers such as Internet Explorer or Firefox, web services enabled custom applications such as Spatial Datamart of the Victorian Government.

Based on the above architecture grids will be managed in the form of master polygon datasets, with each grid polygon assigned a unique identifier. Actual sample data will be held in related textual tables and related to the location for that value (a grid polygon) by the unique identifier.

PEMS will not be a source metadata repository as jurisdictions will already have in place standard metadata for their spatial datasets, to the current ANZLIC or future ISO 19115 Australian profile

standards. Therefore PEMS holds only PEMS-specific metadata and that a link will be included to an external metadata repository for viewing and potentially searching purposes.

Versioning is also an important concept. The PEMS data model will address the issue of data versioning, both in terms of a newer version of the published datasets and time series of particular datasets. An example of the latter is Normalised Difference Vegetation Index (NDVI) grid data generated for the national use case sponsored by Geoscience Australia.

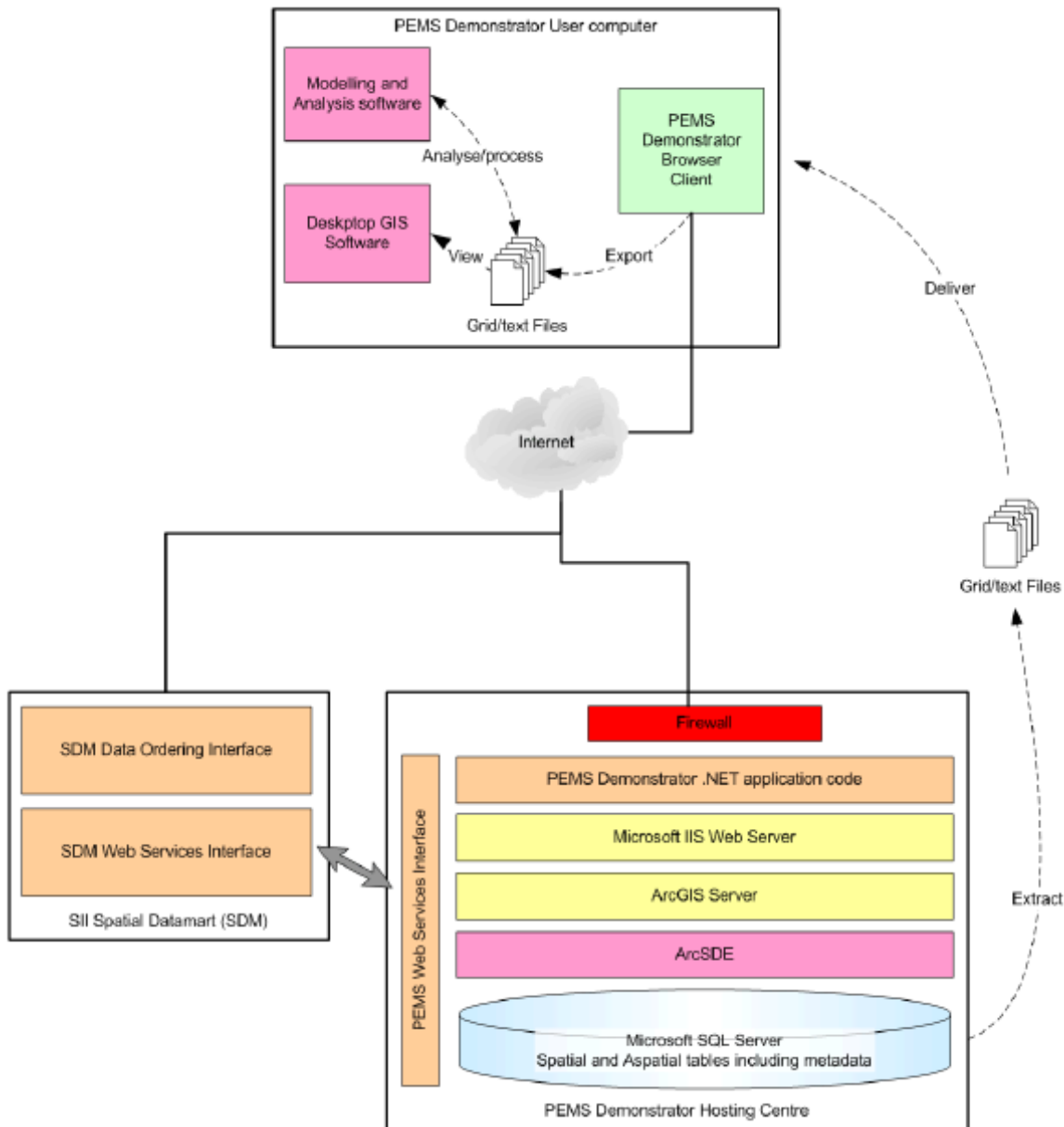


Figure 4. High Level PEMS Demonstrator Architecture

As part of our commitment to meeting ASIBA's expectation when they offered their support, PEMS is designed with interoperability in mind, done through web services standards. To demonstrate this capability, PEMS will provide a web-services interface to enable it to act as a back-end grid data store and engine for other applications. This capability will initially be developed to support the Spatial Datamart (SDM) spatial data ordering application developed by the Victorian Government. PEMS will provide services:

- to report on the available datasets and the areas, formats and projections in which they can be extracted (although this will only be implemented manually in the Demonstrator such that new datasets added to PEMS will not automatically be known about by SDM)
- to enable an extraction order to be validated and placed
- to report back to the requesting application that an extraction order has been completed, providing a URI at which the extracted data can be accessed.

The Demonstrator had to focus its limited resources on demonstrable functions. In this case they included four business functions: visualisation/mapping, modelling/delivery, search (including order of data)/query and administration. These functions are linked through the PEMS screen interface as illustrated in Figure 5.

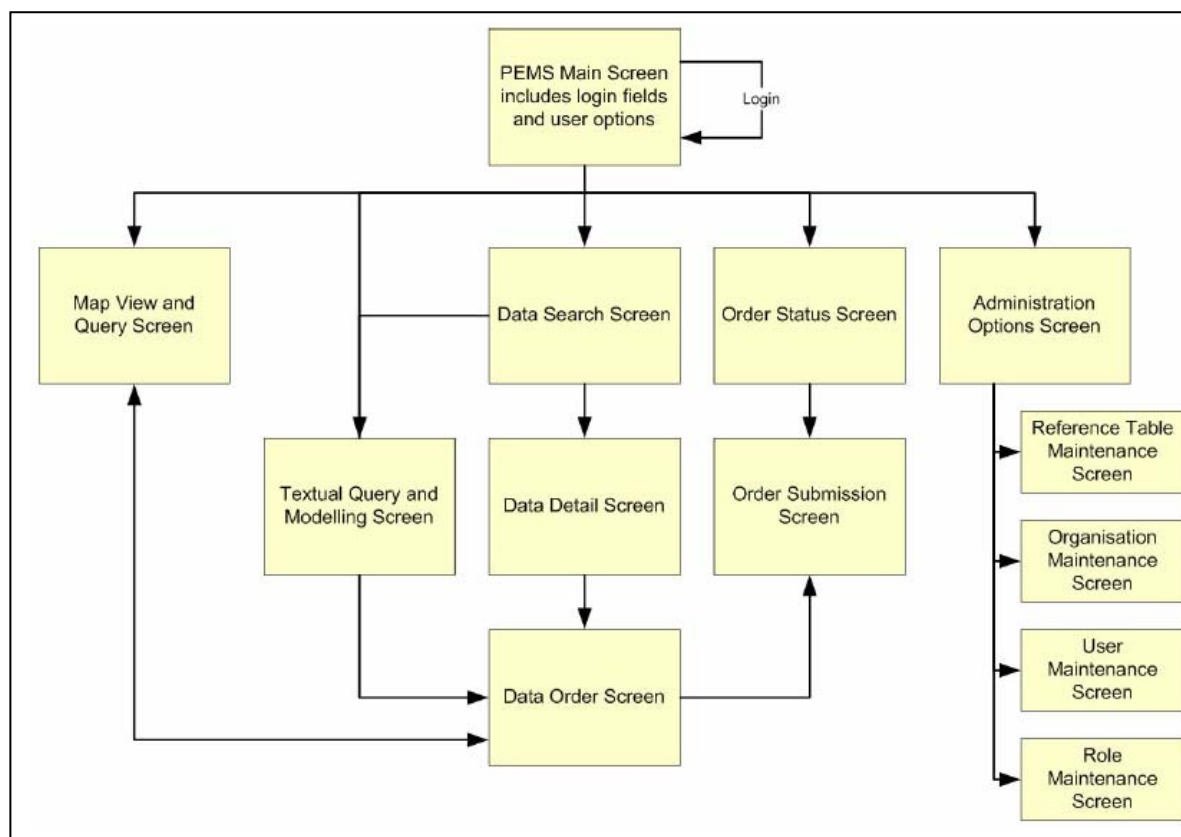


Figure 5. PEMS Screen Flow Overview

Governance

All PEMS datasets are to be under the custodianship of one or more managing organisations and only authorised data custodians/managers/authors/contributors who are given the Data Manager access level, will be allowed to upload and register datasets on behalf of their organisation for specific uses, at least initially. It is expected that, initially, PEMS data (attributes) derived from national and jurisdictional framework datasets will be published for general use when required. As demand for access to PEMS data increases through requests to custodial organisations, each organisation will appoint a data custodian who will work with leaders of data domains, such as land use, soil or biodiversity, in the organisation to publish business datasets for general use. Ideally this will be done in consultation with their counterparts in other PEMS organisations to ensure that business datasets are published in accordance with a set of jurisdictional and even national standards developed for different domains. The rules that guide the processes of consultation and decision making by the data custodians and domain data experts through various communities of practices will form the governance around the registration and publication of PEMS data for use by the wider communities.

Note that it is a requirement that datasets derived within PEMS may subsequently be registered as normal PEMS datasets to enable others to access them. The role required for deriving a dataset by some query or modelling process must be different to that required to register the resulting dataset for general access ie some users may be able to derive datasets for their own use only and would need to have a suitable authorised person register them for general access.

Arguably, the registration processes represent a set of critical gate-keeping PEMS processes that determine the contents of PEMS, which will, in turn, determine the perceived value of PEMS to potential adopters. To instil confidence in PEMS, the technical tools for registration must

complement the governance arrangements designed to encourage data uploaded by individuals into PEMS for specific applications while ensuring that the data eventually published on PEMS for general use is appropriate for that purpose, both from the points of view of data users and custodians.

PEMS Administrator(s) will be provided with screens to create and maintain organisations, users and roles and to assign permissions to roles. Data Managers will set access rights for roles to datasets as part of the metadata entered at upload time and modified thereafter. Anonymous access will be allowed to general functions such as the mapping screen and possibly extraction of publicly available datasets. This will be on the basis of a "public" role which all users, anonymous and registered, will belong to. Roles will be created by the Administrators and will be used to group particular access levels to a set of datasets for assignment to one or more users.

Innovations

The PEMS Demonstrator has adopted an innovative approach to developing a grid cell data infrastructure. Some of the innovations have been described in the previous four subsections. In short the innovations include:

- Separating the publication data store from the storage data store with the former storing grid cell data in generic aspatial data tables for easy access and use
- Developing a first version of the national nested grid as a framework for managing grid cell data in a consistent and repeatable manner across Australia and its jurisdictions
- Using a set of identifiers to link the alphanumeric publication data store to a spatial system; while the former can support basic spatial query, simply modelling and data order services, the latter can provide more advanced spatial visualisation services critical in spatial applications nowadays
- Developing a data model that manages versioning of data, including time series data
- Designing the system to support governance
- Using the CRCIS collaborative framework to build and promote the technology

Implementation of the PEMS Demonstrator

By February 2008 the PEMS Demonstrator would have developed an agreed set of functions selected from its vision for the production version. These functions support four use cases sponsored by the core participants and three demonstration scenarios. Not all functions will be automated; some will be mocked-up to show the intentions and linkages with other functions. The performance, scalability, flexibility and robustness of the system will be at the basic level, as expected of a demonstrator.

The four use cases are listed below and described in the sections that follow.

- National seasonal crop monitoring and forecasting
- Develop and demonstrate a market-based approach to environmental policy on private land
- Wildfire planning: Consequence of Loss modelling
- Land use data, modelling and reporting.

National seasonal crop monitoring and forecasting

The main objective of this use case is to enable the monitoring and modelling of spatio-temporal changes in a pilot study area covering the Mallee area in Victoria using time-series remote sensing data, within the PEMS environment. This use case would also establish the:

- value of PEMS for monitoring temporal changes in landscape at regional scales;
- capability to provide agricultural crop condition information on a periodic basis; and
- standards for a hierarchical, national grid that enables multi-scale data analysis.

For the pilot study area chosen for this use case, two forms of Vegetation Indices (VI), Normalised Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) serve as inputs for modelling and monitoring spatio-temporal changes within PEMS. Time series NDVI and EVI images derived from the satellite data and conforming to multiple resolutions of the hierarchical

grid will be input into PEMS. The PEMS environment would enable the querying of these time series VI data sets based on user specified geographical extent or point location in order to extract information about vegetation condition. The onset of greenness, peak greenness and senescence which are useful indicators of vegetation condition or drought, can be determined from examining the temporal profiles of VI for any location within the PEMS grid (Figure 6).

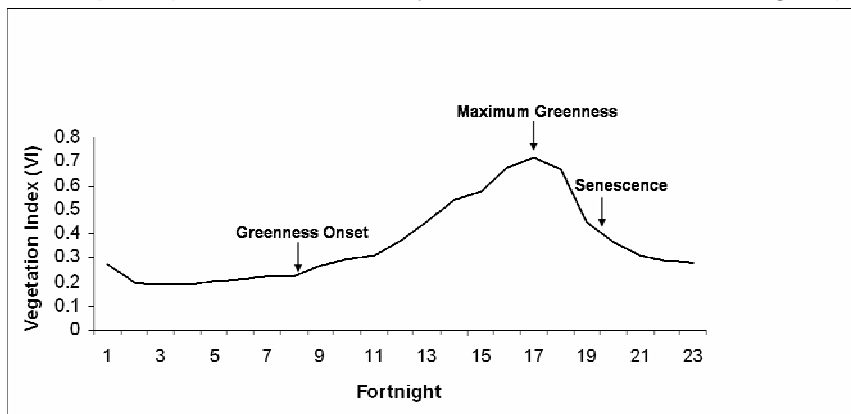


Figure 6. Vegetation Index profile showing key phenological stages

Comparison of the current crop season's VI with a reference (normal crop season) or the long-term average VI could determine whether a crop season is expected to be below normal, above normal or normal for one or more grid-cells or pre-defined administrative units (e.g. Statistical Local Area or Statistical Division) within PEMS. Significant shifts in the VI trends over time or inter-annual changes in land use / land cover can be identified through analysis within PEMS.

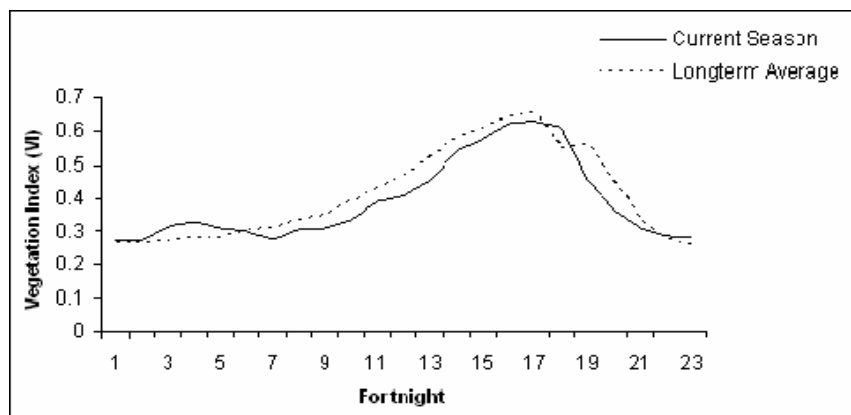


Figure 7. Vegetation Index profile for current season crop compared to the long-term average

Basic VI statistics by location or land cover type queried using PEMS (e.g. mean, standard deviation or histogram of VI for a user selected area) can be output as a chart or simple text. PEMS would also enable queries over multiple grids to compare VI statistics for data sets with different spatial resolutions. Trend analysis of the time series VI data within PEMS would produce charts showing departure from long-term average for the selected location or cover type, similar to the chart shown in Figure 7. The ability to export query results as GIS data would also enable further analysis of the results within a GIS environment.

Develop and demonstrate a market-based approach to environmental policy on private land

The "Market-based solutions to redress landscape decline" (MBS) is a Victorian Government project to develop and trial one comprehensive set of policy mechanisms/instruments. The mechanisms aim to improve governments' ability to procure environmental services from private landholders. The approaches are based in part on a pilot project called "EcoTender". EcoTender uses a model called Catchment Management Framework (CMF), which, in turn, builds upon a ground water model called Catchment Analysis Toolkit (CAT), to assess the value of environmental services that can be offered by a land manager/ farmer as part of a special tendering process (Eigenraam et al. in press).

PEMS will provide a standard data environment to manage the grid cell data required by CMF and CAT in a consistent format, and integrate with an existing spatial data ordering application called Spatial Datamart in Victorian Government to provide the web-based order and delivery of the above data in a standardised format. Re-sampling of data to a different resolution is also possible. The PEMS standard data environment also will allow the easy publication of outputs from models such as CMF and CAT so that new data can be viewed together with existing data to better inform decision making at the farm-scale enterprises level through to State-wide initiatives.

Wildfire planning: Consequence of Loss modelling

As part of its wildfire risk management model the Victorian Government has mapped the value of a wide range of assets across the state to provide a valuable starting point from which to assess the total Consequence of Loss of an asset due to fire. However, the relative values of an asset may differ at state, regional and local levels; particularly when planning for fire management at these different levels.

It is anticipated that PEMS will provide an environment where the asset value data is accessible and available in a standardized, interactive, visually rich format to allow local fire planners to apply weightings to asset categories or, where there is no damage but a disruptive effect to an asset, a disruption value. This allows improved community engagement significant in fire management by capturing local knowledge and providing a local representation of the loss of an asset.

Land use data, modelling and reporting

The Victorian Government is interested in developing spatial data infrastructure to support the better understanding of the impacts of current and future land use practices, by means of a methodology that spans across the data, modelling and reporting continuum (Figure 8).

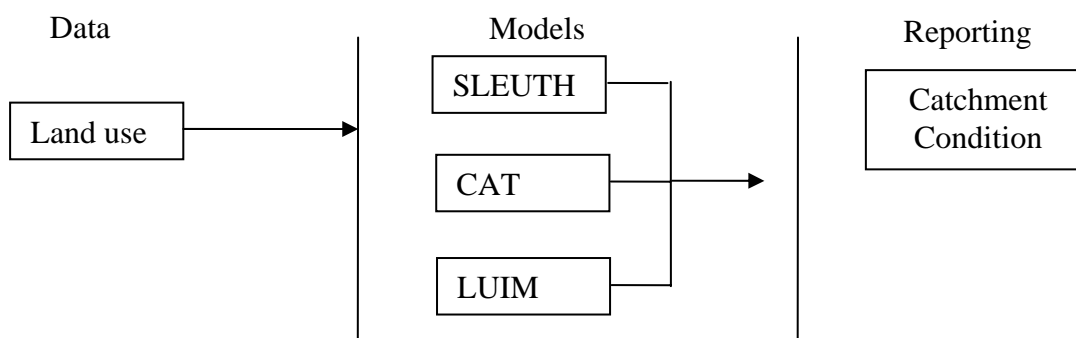


Figure 8. Land use data, modelling and reporting methodology

It is envisaged PEMS will provide a data management environment supporting data versioning and associated metadata, which allows users to specify and select individual land use classes or groupings for further analysis, processing or download. PEMS will be required to support model metadata ranging from what data a model requires for execution, through recording details of model runs associated with data stored in PEMS, to ultimately supporting an eScience style implementation (Pettit 2008) for model execution, chaining, coupling and delivery. PEMS will be evaluated for its ability to accurately convert and store an instance of social/economic/environmental indicators and to support “just-in-time” policy and decision-makers requests for information, including time series analyses.

Demonstration scenarios

There are three demonstration scenarios to showcase some generic capabilities of PEMS not evaluated through the use cases.

1. National application example: demonstrate the management of national datasets at 10km resolution, possibly climate data such as mean monthly temperatures and annual rainfall.
2. Cross grid system example: demonstrate cross grid system data re-sampling using available land use data at 100m (4 second) and 20m (1 second) grids.

3. Site suitability example: demonstrate site suitability analysis by means of sieve mapping using PEMS query functions with several 20m datasets for selected local government areas.

PEMS is also working collaborative with the Victorian eResearch Strategic Initiative to demonstrate how an eResearch platform, namely, Ecoinformatics, can support integrated climate change adaptation research. This is done through the adoption of the proposed national nested grid developed for PEMS as the basis to support data transfer across climate change models.

Values

The main value of PEMS as a grid cell component of the ASDI lies in the typical cost structure of developing GIS applications. According to a study conducted by the Federal Geographic Data Committee (FGDC) of USA (2000) about 75% of the costs are expended in locating, assembling data while only about 18% and 7% in application development and actual productive operations, respectively. Therefore, it is imperative for governments to facilitate the documentation, discovery and access to spatial information assets, both new and existing, within their jurisdictions so that these assets can be used effectively and efficiently to address the increasing complex social, economic and environmental challenges the world is facing.

In the context of the life cycle of data utilisation described earlier and the FGDC finding, the steps of *search/access*, *process*, *integrate* and, to a certain extent, *design* constitute about 75% of the cost of typical environmental modelling/analytical work. The steps of *analyse* and *report* only constitute 25%. These do not include the expenses of managing the data to deal with data requests from other users later on, ie., *publish*. It is envisaged that PEMS can support the steps of *search/access*, *process*, *integrate*, *publish*. To a certain extent, PEMS can also support *design* by providing information on the attributes available in its data store. It can also support *analyse* and *report* through simple/customised modelling and routine reporting processes developed on the PEMS platform. As a managed service, PEMS is expected to be able to make these steps more efficient through reducing duplication and automating the functions underpinning the steps.

More importantly, PEMS also offers strategic value to national and jurisdictional environmental management. The proposed national nested grid and the two-tier data management approach provide a useful framework, within which governance could be developed by data domain experts to standardise development and publication of commonly used environmental attributes. This will ensure their fitness-for-purpose and future reuse. It can also facilitate integration with compliant data nodes elsewhere. By developing functions around the framework and the associated governance, PEMS provides a managed environment for current data of known and consistent quality, which can be integrated with data from other sources to support ongoing environment modelling and analyses. As the grid cell data holdings grow, and the very efficient way “spatial” queries by business managers and decision makers can be supported through simple SQL queries, PEMS is potentially a corporate resource to support routine spatial queries by all.

Conclusions

Scientists in Australia, including environmental modellers and GIS analysts are playing an increasing role in providing the evidence needed for informed decision making and policy development in response to climate change. They need grid cell data for their modelling and analyses. Based on a typical data utilisation life cycle for these scientists, it is argued that they need data management support that extends beyond the Australian Spatial Data Directory and covers the steps of *search/access*, *process*, *integrate*, and *publish*.

The Platform for Environmental Modelling Support (PEMS) Demonstrator project, sponsored by the CRCSI and supported by ANZLIC and ASIBA, is a prototype that can provide the extended data management support identified. PEMS is designed as a grid cell data component of the Australian Spatial Data Infrastructure that incorporates a number of innovations such as a national nested grid system and a two-tier data storage model. The four components of PEMS: data, standards, technology and governance, are described together with the four use cases and three demonstration scenarios designed to evaluate its value. Apart from the usual cost-efficiency gains

as identified in the Federal Geographic Data Committee study in 2000, strategic benefits that PEMS can offer include the provision of a framework that provides access to current, consistent and authoritative data; facilitates data domain standardisation and inter- and intra-jurisdictional data integration.

Acknowledgments

This development work is being conducted within the CRC for Spatial Information, established and supported under the Australian Government's Cooperative Research Centres Programme. The views expressed in this paper are those of the authors and do not represent the views of their organisations.

References

- Annoni, A. and Craglia, M. (2005) *Towards a Directive Establishing an Infrastructure for Spatial Information in Europe (INSPIRE)*, paper presented at FIG Working Week 2005 and GSDI-8, Cairo, Egypt April 16-21. Retrieved on 28 April 2007, from <http://www.gsdi.org/gsdiconfproceedings/gsdi-9.asp>
- ASDD (2008) Australian Spatial Data Directory home page <http://asdd.ga.gov.au/asdd/>, accessed on 10/01/08
- Cooperative Research Centre for Spatial Information (2007) *Use of Raster Data – Summary*, paper published at the PEMS home page. Retrieved on 30 December 2007, from <http://www.spatialvision.com.au/html/pems-project.htm>.
- Eigenraam, M., Strappazon, L., Lansdell, N., Beverly, C. and Stoneham, G. (in press) Designing frameworks to deliver unknown information to support market based instruments, in Otsuka, K. and Kalirajan K. (Eds.) *Contributions of Agricultural Economics to Critical Policy Issues*, Blackwell Malden, MA.
- Federal Geographic Data Committee (FGDC) (2000) *Financing the NSDI: National Spatial Data Infrastructure - Aligning Federal and Non-Federal Investments in Spatial Data, Decision Support and Information Resources, Revision 2.0 for public comments*, p.37. 232 pages. Retrieved on 30 December 2007, from <http://www.fgdc.gov/library/whitepapers-reports/>.
- IPCC (2008) IPCC home page <http://www.ipcc.ch/>, accessed on 10/01/08
- Nebert, D. (2004) Ed. *Developing Spatial Data Infrastructures: The SDI Cookbook*. Version 2.0, GSDI. Retrieved on 28 April 2007, from <http://www.gsdi.org/gsdicookbookindex.asp>.
- NDN (2004) *NDN05-02: In brief: the National Data Network*, published at the National Data Network web site, retrieved on 18/04/08 from <http://www.nationaldatanetwork.org/>
- Onsrud, H. (1999) *Status of NSDI around the World*. A report on the status of SDI Initiatives around the world, University of Maine, Maine.
- Pettit, C. 2008 An eScience platform to support climate change adaptation in Victoria, presented at American Associate of Geographers Annual Meeting, Boston Massachusetts, 15-19th April 2008.
- Stern, N. (2006) *Stern Review executive summary*, published at the British Government Treasury web site, retrieved on 10/01/08 from http://www.hm-treasury.gov.uk/media/4/3/Executive_Summary.pdf.

Appendix 1. Table showing the implications of distance and area measurements for grids stored in geographicals system

Grid Cell Size	Notation		Grid Cell Size			Distance at Latitude (m)					Magnitude (m)	Example Data
	Decimal	Fraction	Degrees (°)	Minutes (')	Seconds (")	0°	10°S	23°26'S	45°S	55°S		
Degrees (°)	1		-	60	3,600	111,120	109,432	101,950	78,573	63,735	-	
	0.5	1/2	-	30	1,800						-	
	0.1	1/10	-	6	360	11,112	10,943	10,195	7,857	6,374	-	
	0.01	1/100	-	0.6	36						-	
	0.001	1/1,000	-	0.06	3.6						-	
	0.0001	1/10,000	-	0.006	0.36						-	
Minutes (')	30		0.5	-	1,800	55,560	54,716	50,975	39,287	31,868	10,000's	
	5		0.08333~	-	300						1,000's	
	1		0.01666~	-	60	1,852	1,824	1,699	1,340	1,062		
	0.5	1/2	0.00833~	-	30						100's	
	0.1	1/10	0.00166~	-	6							
Seconds (")	30		0.00833~	0.5	-	926	912	850	655	531		
	9		0.0025	0.15	-	278	273	255	196	159		9" DEM
	5		0.00138~	0.08333~	-							
	1		0.00027~	0.01666~	-	30.9	30.4	28.3	21.8	17.7	10's	Landsat ETM
	0.5	1/2	0.000138~	0.00833~	-							
	0.1	1/10	0.000027~	0.00166~	-	3.087	3.042	2.830	2.184	1.770	1's	