Changes in our Spatial Data Delivery

The Department of Lands (Lands) is responsible for the ongoing maintenance of the digital cadastral database (DCDB), the digital topographic database (DTDB), survey control information and various other land-related spatial products (eg. aerial photography and digital elevation models) for New South Wales. Your organisation has previously received (spatial) data supplies and periodic updates from Lands in a format compatible with your GIS/Mapping systems.

Over the last two years, Lands has engaged in a major systems re-engineering process to integrate spatial data using contemporary database technologies. Additionally, we have re-defined our geospatial product/tools mix to enable better customer service. We are using our new toolset to rebuild our maintenance and delivery environments in a manner that will provide better quality control and information tracking.

Earlier this year, we conducted a customer survey aimed at determining ways in which we could better serve our clients in the maintenance and delivery of DCDB information. Many of the changes described below are a result of feedback from that survey.

System changes

Significant changes have been made to the way that we model and store spatial data. These changes will affect the way you receive data as well as the actual data content. This bulletin is provided to enable your organisation to assess the impact of these changes on your system and minimise any disruption.

For customers anticipating an annual update, there will be a final data extraction that will provide update information as at close of business Friday 9 July from the existing Genamap-based maintenance repository. This update will be supplied to you in the same format that your previous update information has been supplied (ie. having due regard to area of interest, datum, projection, interchange standard and media used).

From 9 July 2004, all maintenance of the DCDB was suspended in order to facilitate the migration of data into the new spatial database environment. During this suspension, Lands spatial data maintenance staff will undergo training in the new system, applications and workflow practices to ensure the resumption of maintenance with minimum delay.

Maintenance is planned to recommence in early September with the database back to a fully maintained state by early December 2004. It is likely that delivery of update information (from the new system) will not be available until this time. There will be a period of ‘beta testing’ of the new delivery system before it is made available on a production level.

While we understand that this delay may cause a disruption to your system, during this period Lands will be assisting clients in adapting to the new processes required to changeover to the new system.

Characteristics of the new delivery system

The new system will provide our customers with ‘on-demand’ or ‘set and forget’ data delivery processes. This will mean that you will be able to manage your own data delivery frequency and associated parameters in a manner that best suits your organisation. Data will be transmitted to your organisation using a variety of techniques including email, FTP, download to media requests, synchronous and asynchronous HTTP transfer.

Baseline data re-supplies

Data structure changes

There will be significant changes in both the data model/schema used for the storage of data within Lands and also in the granularity and density of some of the data features. Lands is re-processing all of the spatial data as it is migrated from the legacy Genamap system into the new database environment.
The re-processing includes:

- removal of many unnecessary vertices from feature chains (the substantive shape of the feature will not be impacted);
- the aggregation of individual geometry polygons into parcels (previous supplies often provided several polygon geometries that required processing based upon a common ‘tag’ identifier);
- the aggregation of polygons into lots. All lots that are composed of multiple polygons will be aggregated under the same ‘cadID’ in the form of a multi-part polygon;
- the rationalisation of the current land unit identifier (‘luid’) polygon attribute into a schema that is unique across all features classes within the DCDB; and
- many undesirable artefacts of the existing datasets will be further cleaned/removed from the data (eg. topological inconsistencies).

Within the new environment, data is broken out into a more granular form and additional classes of data have been established. Of particular note is the clear distinction between the legal cadastral fabric and the road/rail corridor data. Previously, clients were left to extract their own corridor data based upon the ‘tag’ information provided (eg. 023/028 prefixes). Lands is establishing a separate ‘corridor’ class that will ultimately reflect a complete extraction (and projection) of the road corridor network. An additional road centreline network is also being established as the data is migrated.

The most obvious impact of these changes is that previously supplied base data will need to be updated in order to remain in synch with the new system. Clients may of course continue to run with their existing datasets, but their ability to implement a continuing update process based upon existing practices will be significantly impaired. It is strongly recommended that your organisation consider receiving a complete refresh of the data covering your area of interest with a new baseline that is consistent with Lands’ data holdings. This will be an essential requirement if you are to take advantage of the newly developed ‘hands off’ data supply service that will be provided to the community during 2004-2005.

Supported coordinate systems

For previous data supplies, Lands has provided data in a variety of coordinate systems on several datum. In line with government data integration policies, Lands will only be supplying data from the new delivery environment in the officially endorsed GDA latitude, longitude coordinate system or the equivalent Map Grid Australia (MGA) values. If you are awaiting an annual update, the data you receive will be current as at 9 July 2004. This data will be delivered in the format (and coordinate system) that you are familiar with, however subsequent updates will be supplied in the new delivery format.

Supported formats

Baseline supplies and subsequent updates will be available in the following supported formats.

<table>
<thead>
<tr>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WKT</td>
<td>OGIS well known text</td>
</tr>
<tr>
<td>WKB</td>
<td>OGIS well known binary</td>
</tr>
<tr>
<td>MID/MIF</td>
<td>MapInfo transfer format</td>
</tr>
<tr>
<td>Shape</td>
<td>ESRI shapefile format</td>
</tr>
<tr>
<td>SDO</td>
<td>Oracle database SDO_GEOMETRY (ASCII representation)</td>
</tr>
<tr>
<td>SP8</td>
<td>Informix database Spatial 8.xx geometries (ASCII representation)</td>
</tr>
</tbody>
</table>

What are the benefits of receiving a new baseline to my organisation?

In recent months, Lands has visited a cross-section of organisations (thirteen from the three levels of Government, including utilities) that frequent recipients of update data. The majority of these organisations were themselves undergoing a process of mapping/spatial support system renewal. It is likely that this is being driven by the need to deliver products and services (both to external and internal clients) in a more timely and cost effective manner. Most of these clients have seen immediate benefits in adopting new processes and models for the ongoing maintenance of their primary mapping base.

Key benefits include:

- lowered ongoing maintenance costs;
- ability to receive online incremental updates;
- spatial data synchronised and current with Lands reference set;
- fail-safe acquisition of specified Lands data;
- extensible to other feature sets/layers/classes of data available from Lands; and
- web based administration.

Ten of the interviewed clients visited were keen to realise the benefits of an automated data update service as soon as possible.

Data Updates

Online updates from Lands’ spatial datasets will be available following the initial baseline supply. Irrespective of the preferred physical transfer method, all data will be transmitted in one of two types of transfer bundle.

UPDATE bundles

This type of data bundle will occur as a result of normal
everyday maintenance activities against the parcel base (eg. consolidation or subdivision activity, attribute changes or update etc). An update bundle will generally result in no movement of the underlying parcel base since new (or modified) geometries are made to fit coincident with the pre-existing cadastral fabric. This is largely synonymous with the existing practices that have been applied in the Genamap environment. These bundles are generated on demand and contain all changes that have occurred in your area of interest from the date of last supply of updates to the date of request.

**UPGRADE bundles**

This type of data bundle occurs when we apply a large-scale adjustment over a given area of the cadastral fabric. This will normally occur when there is sufficient data of a higher order of precision and/or accuracy in an area to warrant the application of a block adjustment. The result of this adjustment process will be that the geometrical shape and position of the parcels in question will change.

All data transfer bundles transmitted by Lands will be:

- serialised;
- in XML (eXtensible Markup Language);
- in GDA latitude, longitude coordinates or MGA coordinates; and
- predicated upon the client having previously received and loaded a new baseline dataset.

In the case of upgrade bundles, a transformation grid will be supplied in the bundle to enable you to apply a shift algorithm to localised data specific to your organisation that may be affected by the upgrade shifts. The transformation grid parameters provided will enable your organisation to utilise the publicly available NTv2 grid shift algorithm to shift any localised data elements. Processing should always cease where a discrepancy is discovered between your data set and the update bundle being processed and remedial actions taken.

**Serialised data**

Each data feature supplied will be serialised with a unique identification code (eg, a cadID for cadastral data and a topoID for topographic data). To ensure that your datasets reflect the most current state of the DCDB (or DTDB), you must apply the data bundles in a sequential fashion. Each data bundle will have a sequence number that must be applied to your dataset in the same order as received. This is particularly true in the case of update bundles as they identify new parcels/elements to be added to your data set (new cadIDs), and also identify existing parcel/elements to be retired/deleted from your data set (existing cadIDs).

**XML bundles**

The new form of data delivery will in all cases be ‘self describing’ XML bundles. Each bundle will effectively contain a number of data ‘sections’. In the case of an update bundle there will be meta, insert, retire and lookup sections, whilst for upgrade bundles there will be meta, replace and grid sections.

The META data section describes the known client parameters, the data class features, the coordinate system used, the table and class/attribute mappings and other information required to control the load of data to your systems.

An INSERT data section will contain the actual geometries and attributes that need to be loaded (added) into your system. The geometries will be encoded in one of the following formats (as specified by you):

- GML OGIS geographical markup language
- WKT OGIS well known text
- WKB OGIS well known binary
- MID/MIF Mapinfo ASCII transfer format
- Shape ESRI shapefile format
- SDO Oracle database SDO_GEOMETRY (ASCII representation)
- SP8 Informix database Spatial 8.xx geometries (ASCII representation)

Also within the INSERT section will be any new attribute data that needs adding to your system. This will always be in a simple ASCII form and referenced via the class attribute mapping information supplied in the META section.

The RETIRE section will nominally contain a list of feature ID’s (eg cadIDs) that will require either retiring or deleting from your system. It should be noted that Lands maintains a persistent dataset in the delivery environment (no data is ever deleted). This is implemented by using a start date/time and end date/time attribute on every feature. Your organisation may not maintain spatial data in this manner, therefore the way you use the RETIRE data is a design decision.

The LOOKUP section will provide a comma-delimited tabulation of useful information that may assist you in affecting the load of data onto your system. It is known that various clients have used different keys when creating their spatial data sets and that programmatic access methods to locate existing features or parcels will vary considerably. A typical LOOKUP would minimally contain (for each feature referenced in the INSERT and RETIRE sections) the unique feature identifier (cadID), lot, section and plan information in separated out and concatenated
tag form (eg. “18,A,252868, 013/18/A/252868”) and polygon centroid or start vertex coordinate of a feature.

A REPLACE section will have the same physical structure as an INSERT section except that your processing actions should be to replace or modify the equivalent feature in your current data set.

The GRID section will provide you with the parameters of a NTv2 style shift pattern that can be used to adjust your localised spatial/mapping data within the area affected by data upgrade activity.

Our new automated update service

At Lands, we are designing our new online delivery service with our customers in mind. In conjunction with our server-side extraction and delivery processes, two client-side processes are being developed. These will reduce the complexity and technical knowledge required to implement the system. Both processes have been developed using Java (an implicit assumption is that your systems can support a Java Runtime Environment) and offer support for either synchronous (database) or asynchronous (FTP, mail, file based) data transfers.

Type 1 transfer

This is a synchronous data transfer that is supported by a Java servlet that resides behind your firewall. The servlet accepts and processes HTTP or HTTPS transfer requests from the Lands-based Postmaster service. It performs data packet validation and acknowledgement with the Lands Postmaster service, XML bundle unpacking, database insert/retire/delete processing, commit or rollback processing and status reports. The servlet will process the XML transfer bundle via direct access to your target (or staging) database.

Type 2 transfer

This is a Java-based application that will unbundle and prepare a previously received (via FTP, email or file based) XML update/upgrade bundle and place it into a location of your file system to continue processing via internal means. It is effectively an automation of what would otherwise have been a traditional data transfer via hard media (eg. CDROM, DLT etc).

The above applications will be made available in an open source form so that organisations wishing to extend the code base will have the opportunity to do so.

The NSW Single Land Cadastre (SLC)

Lands is currently undertaking a project with Sydney Water Corporation (SWC) to coalesce the two separate spatial cadastral representations of the Sydney metropolitan area. The goal of the SLC is to eliminate duplication of effort and realise significant improvement in the accuracy and consistency of metropolitan data. Clearly, it is more cost effective to maintain a single version of the cadastre for the community. SWC cadastral data has (generally) been developed and maintained from the surveyed pipe reticulation and hence is highly accurate in many areas. Similarly, Lands data has been maintained and refined to ensure attribute correctness and geometrical/topological consistency. The vast majority of Lands data is already coincident with SWC data due to a previous data capture/remediation project between the two organisations. The SLC will assume the attribution consistency and topological correctness of Lands data and the positional accuracy of SWC data.

There will naturally be a number of areas where the Lands cadastral base shifts significantly as a result of the SLC project. These locations are largely confined to the peri-urban areas where Lands source mapping changed from 1:4,000 to 1:25,000 and where SWC have cadastre based upon the pipe-work survey data. Lands is currently identifying those areas that are most affected.

The Sydney metropolitan project is scheduled for initial deployment of data for test purposes in March 2005. Lands intends that the Single Land Cadastre concept will eventually be available on a State-wide basis.

How do I implement this service?

It is a essential that you reload your systems with a baseline of data that is consistent with the data held at Lands. Ideally, certain attribute fields should be implemented in your system to enable identification of records or features that are referenced in the update/upgrade bundles. For some clients, there is an opportunity to effectively replicate the schema used at Lands and extend this with their own attributes. For others, there may be a need to rewrite code to take data from a file location and into their mapping package.

The Lands supplied Java servlet (and Java application) will require an appropriate Java application/servlet runtime environment (preferably in your DMZ) and support for the various communication protocols (HTTP/S, or FTP/ SFTP if file transfer is your preferred method).

Assistance in the specification of an appropriate environment will be provided by Lands.
What are software vendors doing to help facilitate this new system?

Lands has approached several software vendors and each has been keen to offer support in extending the abilities of their product to receive online updates via the proposed deployment method. The vendors see the integration of online data updates as providing a valuable service to their NSW marketplace clients. Software vendors will be included in the beta testing phase of the product and Lands will work closely with those vendors to integrate the client-side processes.

Where does this new service fit in a contemporary N-tiered architecture?

Most organisations have adopted a multi-tiered approach to their systems architecture. This has often resulted in a segmentation of the environment into (typically) a firewall that separates external connectivity from internal networks, and then a further separation of internal networks to DMZ and Production. The Lands receiver servlet/application is intended to sit behind your firewall within your DMZ. The receiver servlet will additionally require database connectivity to a target system in your production environment that is used for mirroring the spatial data. This target may either be a live production system or can be a staging database that is subsequently updated by you onto your production environment. In the case of data supplies that utilise file uploads (eg. FTP) and similar, Lands will supply a Java application that can be used to ‘unpack’ the data bundles that are transmitted. As with your existing system, a further (internal) process will then be required to move the data into your production systems.

Delivery system overview

In order for the new delivery system to work effectively, there is a necessity to refresh the client-side spatial systems with a new data baseline. This will ensure that the remote maintenance service is able to correctly synchronise data against your system/database. Once a new baseline has been established, the remote deployment service will maintain the synchronicity of the client data against the Lands master data by a series of UPDATE and UPGRADE data bundles.
resultant files in a pre-determined location on your system. A further process (supplied by yourselves) would then be required to take this data and install it on your system. The ‘unpacked’ data is equivalent to the data you would normally have received as an update from Lands. The local process would require suspending any further database updates as soon as an error is detected.

(1.5) Internally developed load and QA process.
This process is specific to your systems’ environment. It will perform the necessary data loading conversion of the data as delivered (and unpacked) from Lands’ XML bundles into your system.

(2.1) Lands Informix (Spatial) database
Lands is redeveloping its data delivery/deployment environment. A variety of products are to be used to facilitate delivery requests. The remote maintenance system is one such product being developed to satisfy a market need. The core repository for spatial (and selected non-spatial) data is Informix 9.4 running Spatial 8.2 datablade.

(2.2) Lands Extractor servlet
This process is executed on a daily basis for the processing of update/upgrade extraction requests. The servlet examines each registered client detail set for information relating to area(s) of interest, required feature classes, update frequency, date of last update received, data format and other client defined preferences. Once information about client requirements is determined, the servlet then queries the database and constructs one or more update data bundles ready for transfer by the Postmaster service.

(2.3) Lands Postmaster servlet
The Postmaster service is responsible for delivery of update/upgrade bundles created by the extractor process to the client systems. The Postmaster is able to operate in both asynchronous and synchronous data transfer modes depending upon the client delivery type (eg. Receiver servlet – synchronous transfer, FTP – asynchronous).

Implementation schedule

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 July</td>
<td>Genamap maintenance systems turned off to allow data migration.</td>
</tr>
<tr>
<td></td>
<td>Maintenance suspended.</td>
</tr>
<tr>
<td>End August</td>
<td>Final cut of update data to client base to be supplied out of existing system.</td>
</tr>
<tr>
<td></td>
<td>Data migration completed.</td>
</tr>
<tr>
<td></td>
<td>Maintenance backlog clearing starts.</td>
</tr>
<tr>
<td>End Sept</td>
<td>First beta testers installed for online delivery / process update.</td>
</tr>
<tr>
<td>End Dec</td>
<td>First production roll out of client-side online delivery system.</td>
</tr>
<tr>
<td></td>
<td>Maintenance backlog cleared.</td>
</tr>
</tbody>
</table>

The next step
This bulletin describes how online incremental updates will be delivered for the DCDB. Future development will see updates to the DTDB provided in a similar manner. A separate bulletin will describe those plans in more detail.

We would like you to consider what impact the Lands changes may have on your current environment as it is important that we understand what these changes will mean to your organisation. Lands is seeking your cooperation to assist us in developing a series of technical profiles in regards to our client systems environments. This will enable us to better assess the future challenges for our customers.

We have enclosed a technical questionnaire with this bulletin and would appreciate your time in providing us with a general overview of your IT infrastructure as it currently exists and/or will exist in the near future.

Further information
For further information on the development of online access to the cadastre during 2004-2005, please contact:

AI Garside, Manager, Spatial Data Services
T 02 6332 8349
E alan.garside@lands.nsw.gov.au

Chris White, Business Development Manager (Commercial)
T 02 8258 7453
E christopher.white@lands.nsw.gov.au

Carl Ginger, Business Development Manager (Local Government)
T 02 9228 6032
E carl.ginger@lands.nsw.gov.au

Department of Lands
Head Office
1 Prince Albert Road
Queens Square
SYDNEY NSW 2000
T 61 2 9228 6666
F 61 2 9233 4357
www.lands.nsw.gov.au
© 2004 NSW Department of Lands