Progressive Development of a Digital Cadastral Data Base

THOMPSON, Rodney James

1. SUMMARY

A Digital Cadastral Data Base (DCDB) is a big investment for a jurisdiction tasked with the administration of land boundaries. In the past, the development of such a database produced no real pay-back on investment until many years, and millions of dollars had been committed.

The Land Administration Domain Model (LADM) (ISO-TC211 2012) provides a model of a schema in which a progressive creation and improvement of a DCDB is possible, to allow benefits to be obtained even in the early stages of this effort. It also includes the necessary structure to ensure that a useful historical record of the cadastre can be kept. This paper explores issues to be faced in the development of software based on the LADM, which retains the history of the cadastre, and allows for progressive improvement of the data. The examples of the Netherlands Kadaster and the Queensland (Australia) Department of Natural Resources and Mines are used to evaluate the applicability of the approach.

2. INTRODUCTION

Typically, a DCDB is repository where is stored information extracted from the survey process in a form that allows an overview of the relationship between spatial units. The problem has been that the database structures have been such that only "valid" data can be stored, therefore much manual cleaning and correcting of the data is necessary. This is exacerbated by the fact that a DCDB is of limited usefulness until it is complete.

In the past, it was seen as important that all data to be included in any database should be absolutely clean and correct, with validation filters to prevent "bad" data from being accepted. For example, parcel boundaries should close, without undershoots or overshoots. A series of validation requirements have been specified, and the database management systems have been built around this concept of rejecting bad data. Unfortunately, this puts a giant hurdle in front of any organisation. If the data cannot be entered without being correct, it cannot be made visible to a wider audience, and no pay-back can be obtained. Correcting data on input is a difficult (and therefore expensive) process, and the only eyes on the data are those of the data capture operators.

By contrast, cadastral data which is not necessarily "topologically clean" can be correct and very useful. That is to say, there are many products and services that can be extracted from a cadastral data repository which still contains data which may have impurities. It is important to remember that no spatial data can ever be "correct", and invariably has an intrinsic limit to its accuracy. As an example, although data containing gaps, overshoots and undershoots of very small size may be rejected by the software environments, it would be quite useable for many purposes, provided the relationship between properties is quite clear. For example, various mapping functions, including Web Map Services (WMS), Web Feature Services

(WFS) and ad-hoc cadastral maps, searches etc, may be quite adequately provided by such a product.

Ideally, a progressive capture and "cleaning" process should be adopted. The LADM (ISO-TC211 2012) allows for this progressive development by defining 5 levels of encoding:

1. "Text-Based" Spatial Unit
2. "Point-Based" Spatial Unit
3. "Line-Based" Spatial Unit
4. "Polygon-Based" Spatial Unit
5. "Topology-Based" Spatial Unit
   (With "Sketch-Based" as a sub category of Text-Based).

As a DCDB matures, it can be expected that its quality is improved, both in terms of its accuracy, and in terms of its topological purity. This may also involve changes in the level of encoding. For example, a jurisdiction with polygon-based encoding may convert to a topology based form. In addition, the LADM provides for the inclusion of 3D spatial units, which may be intermixed freely with earlier 2D parcels.

It is an important part of any Cadastral database (though sometimes overlooked) that a history of the use of land be maintained. In the days of local maps printed on linen, and working maps marked up manually with updates and changes, there was a form of historic record of the cadastre. This should not be lost in the move to "digital".

In providing this functioinality the Queensland Government, like the Netherlands Kadaster, adopted forms of what is now known as the "Versioned Object" pattern. The LADM also uses this pattern, permitting a permanent but efficient storage of cadastral history within the database itself. In a progressively developed database with history, it must be recognised that older historic data will usually be at a lower state of accuracy and topological purity than has been achieved later. It is however commonly accepted that history of "the cadastre as we knew it" is a valuable resource. In the Queensland case, a view of the DCDB from the 1990's is likely to contain mismatched boundaries, which have been corrected progressively over the years. Thus levels of encoding, topology and accuracy may be non-homogeneous in space and in time. One important issue with history is that it must not be necessary to jettison many years of history if the level of encoding is changed, or to partition the database into incompatible layers to allow progressive improvement.

Rather than each cadastral jurisdiction developing its own database structure based on basic geometric primitives, there is a case to be made for the development of a "Cadastral Toolkit", based on the LADM, which can support all levels of encoding, variable accuracy and topological purity, while maintaining a full history. It would have to allow data quality to vary by geographic and temporal location. This would need to be configurable to allow for country profiles as permitted under ISO19152, thus permitting local terminology and language to be retained. Many jurisdictions are having extreme difficulties in successfully developing database software, so an open source type of development may be indicated.

The paper is structured as follows: Section 3 discusses errors and validation in a spatial database, Section 4 adds the context of the LADM Section 5 discusses history within the model, Section 6 proposes a data model, and Section 7 summarises and suggests further work.

3. ERRORS IN SPATIAL DATA

In building any database, it is clearly desirable to keep the data as correct as possible. This applies to non-spatial as well as spatial databases, and in all database, there are various types of errors that can be:

Corrected automatically. E.g. "sURNAME" can be corrected to "Surname"
Detected and rejected, allowing for manual correction. E.g. "1O0" as a number.
Passed undetected. E.g "Smith" instead of "Jones".

In addition to the desire for correct data, most of the mainstream spatial Database Management Systems (DBMS) define validation checks that are applied to incoming data to ensure removal of data that will cause malfunctioning of application software. For example, if software is relying on areas of parcels, then a parcel which does not form a valid cycle may cause a crash, or even worse, an incorrect result.

3.1. Validation Failures

There are many types of problem in incoming data that can prevent acceptance into a mainstream spatial DBMS. This is not an exhaustive list, but is intended to indicate classes of issues. Undershoots, overshoots, knots and other errors of small magnitude are discussed in section 6.2.2.6 of ISO19107 (ISO-TC211 2003), which categorises such data as "spaghetti". Some of these errors can be taken out automatically, but not all. Certainly, the simple "snapping" of near points has not proven satisfactory, because such an action can cause a loss of information (for example, see Appendix I). As ISO19107 says: "The real problem with 'spaghetti' comes in that the heuristics (either manual or automated) used to correct the problems often result in additional, but different factual errors. This can be a severe quality issue for geometry." (ISO-TC211 2003 page 27).
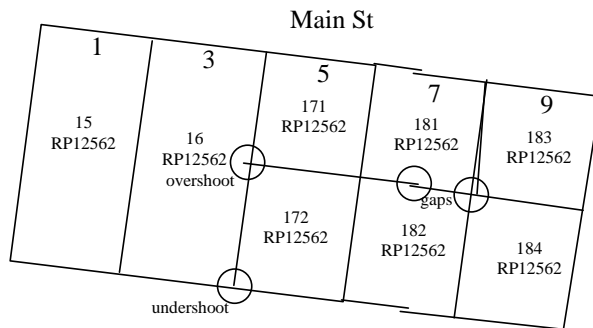

Fig 1 Typical validation errors, gaps, overshoots and overshoots

Some of these errors originate in older manual digitising processes, some in edge-matching of original source documents (Fig 1) and some (such as knots – see Fig 2) due to algorithmic correction of other errors. As a general rule, only very small errors can be successfully corrected algorithmically, and the problem of determining tolerance values should not be under-estimated. In 3D cadastre, there are even more restrictions on automatic error correction (Thompson and van Oosterom 2012).
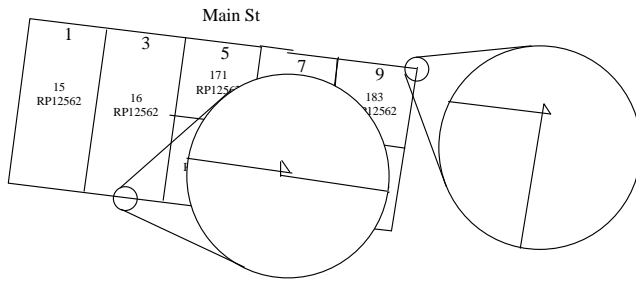
Fig 2 Knots: Left – digitising error, right – end loop.

3.2. Non-Validation Errors

There is a large class of errors that are not detected by validation. For example, the simple misplacement of a vertex, not causing any change in the topology of the parcel fabric, or small "kinks" in lines which do not create a knot.
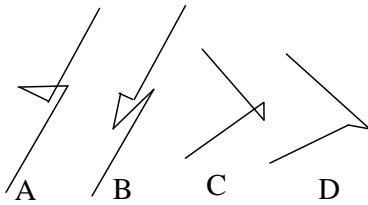

Fig 3 A: a knot which raises an exception, B: a kink which does not, C: A line-joining artefact which raises an exception, D: one which does not.

It has been observed that whenever data is transferred from one software environment to another, the receiving environment will find errors that the other has not found. This has been characterised as superiority of the receiving system. ("System B is better than A because it finds errors A has missed"). This is a misreading of the situation, and should really be characterised as a fault in both system. Specifically, a database cannot rely on incoming data which has been validated by software from a different manufacturer. It must re-validate it to its own requirements.

3.3. Incorrect Encoding

Other more serious errors include the mislabelling of polygons (e.g. 22 instead of 2), miss-attributing geometric objects, etc.; which may not be detectable by validation, or may only be caught by "reasonability checks".

3.4. Reasonability Checking

Various more or less sophisticated forms of reasonability checking are possible, but can be difficult to maintain. All of the checks described below may accidently reject valid data:

> Incorrect Positioning of Vertices: May produce an obvious error which does not cause validation failure.

> Knots: Very small lines that do not show at map scale may indicate knots or kinks.

> Gaps and Overlaps: Close parallel or near parallel lines may indicate gaps or overlaps.

Incorrect Encoding: In Queensland, a lot may consist of multiple spatial units. Miscoding of attributes can result in a lot comprised of spatial units kilometres apart.

4. THE LAND ADMINISTRATION DOMAIN MODEL

It might be thought that the levels of encoding defined in the LADM were intended to be separate, and not mixed in the one database, but this would an unnecessary restriction. There may be major advantages in creating a database and maintenance structure that will allow a Cadastral jurisdiction to progress up through levels of maturity at a pace they can afford. As has been recognised by in the "Social Tenure Domain Model" (SDTM) (Lemmen 2010), it is apparent that there can be great value in a data repository which is of varying levels of accuracy and of quite basic encoding level. It will be suggested here that this principle applies to cadastral databases in general.

4.1. Encodings:

The Land Administration Domain Model (LADM) (ISO-TC211 2012) defines 5 levels of encoding of the geometry, based on the level of maturity:

Text Based Encoding

In this encoding, there is no stored geometric data, and the spatial unit is defined with descriptive text, perhaps accompanied by various diagrams and images. This approach can rigorously define the cadastre, but very little computational analysis is possible.

Point Based Encoding

Fig 4 Point-based encoding against a geo-referenced image (mock-up)

Line Based Encoding

It is relatively cheap to scan and vectorise old cadastral maps, but significantly more expensive and time-consuming to convert the resultant line-work and attributes to a topologically correct encoding. The line-based approach allows a map consisting of labels (property identifiers) and lines to be made available (for example via a Web Map Service). For example, in Fig 5, although there are clearly gaps and overlaps of the sort that would result from very basic scanning, the result would be quite useful for certain purposes.
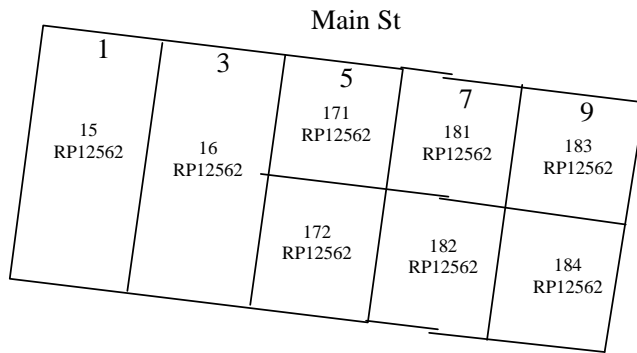
Fig 5 Hypothetical cadastre scanned from existing paper maps

Polygon Based Encoding

This is the path of development adopted by the Queensland Government (Australia). Capture started by digitizing existing paper cadastral working maps, which were then corrected and the identification added. The software available at the time (early 1980's) was limited in capacity, and good topological consistency could be only guaranteed within a limited extent. Therefore the data was divided into "segments" which consisted of up to 300 land parcels. Within individual segments correct joining of parcels was assured, but between segments, very small gaps and/or overlaps of boundaries were common. In building a complete DCDB, the parcels were stored as individual simple polygons, with no attempt to prevent mismatching of boundaries, but the update software was written to ensure that no new mismatches could be introduced. As a result (combined with warning reports), over a period until about 2003, a complete non-overlapping coverage was obtained.

At this level of encoding, all the products available to a line-based encoding are possible, along with the ability to calculate areas, perimeters etc. Once a continuous partition is achieved, all the products of a topology-based approach are also available.

Topology Based Encoding

This is the highest level defined, and provides the functionality of all the lower level encodings. Adjacency is directly encoded n the databanse. This is the path chosen by the Netherlands Kadaster.

Topological breakdown

All levels of encoding 3 or more can exhibit breakdown of the topology. At level 3, this can be the omission of line(s) causing adjacent spatial units to be combined, leading to more than one centroid in a parcel; or an extra line creating a spurious parcel. At level 4, the boundaries of adjacent spatial units may mismatch, leading to gaps or overlaps of neighbouring parcels. At level 5, the encoding of the left or right parcel(s) on a line may be omitted – leading to an unclosed boundary on a spatial unit – particularly a non-base parcel.

4.2. 3D Spatial Units

All the levels of encoding within the LADM are available to be used with 3D cadastre. It might be unusual to consider the lower-level encodings, (such as line-based), but it should be

noted that the commonly used "building unit" form of 3D spatial unit is in effect a text based spatial unit. In this form of spatial unit, the extent is defined by the walls of the building, and often (as in Queensland), are only defined in a form such as "Building Unit Lot 45 on Lot 23 of SP12345". Note that lot 23 is defined geometrically, but the individual building unit is not. There is also a fairly common combined approach – where the floor plan of the unit is defined geometrically, but the only elevation information is textural – such as "on floor 5". The LADM can represent all of these variants and combinations – see Section 6.2.

Three dimensional parcels occur commonly in areas of high property values, and in these areas, the accuracy of survey likewise tends to be higher. Further, it is easier to justify the costs of careful data encoding in these regions. It is easy to envisage a cadastre consisting of a mixture of 2D and 3D parcels using a high level of encoding in the city areas, combined with low-level encoding of lower accuracy information in less dense regions.

4.3. Non-Base Parcels

While the aim of a cadastre is to account for real estate property in a jurisdiction, and commonly this defines a complete and non-overlapping coverage of the area of interest; there are several types of property rights, restrictions and responsibility that are defined as secondary interests to the base coverage. These can be considered to be non-base spatial units. Where the DCDB is being progressively improved in accuracy, the positions of the cadastral points "move" – that is they are redefined, without any change in the topology of the spatial units. Some of the defined constraint regions should move with them and maintain the topological connectivity, while some may not. For example, easements generally are defined in relation to the base cadastre, and therefore move with it. Others such as drainage areas are defined in terms of the relief, and so may not. Difficulties arise where a secondary interest is defined partially in terms of cadastral boundaries, and partially independently, effectively preventing any fully automatic solution.

Universal secondary interests (applying to all spatial units)

These do not need to be recorded in the cadastral database except possibly as a reference to the appropriate legislation. (For example rights to explore for minerals).

Secondary interest as a collection of spatial units

Jurisdictions will group property units for administrative purposes (see Fig 6). Note that there are several variants on the simple hierarchy:

Cadastral Section and Cadastral Municipality identifiers are embedded in the spatial unit identifier, meaning that any reference to a spatial unit carries this information.

The other hierarchies are implemented by encoding an administrative area code in the spatial unit record.

Complex is a partial hierarchy – it only groups spatial units which record apartment rights.

Parish and County are now largely obsolete, only being carried for historical purposes, but do form a complete coverage.
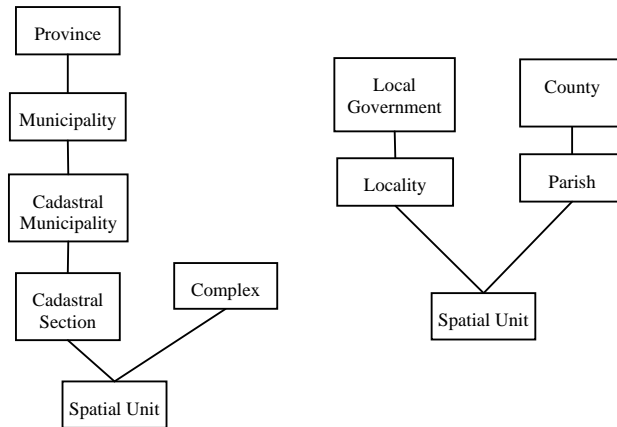
Fig 6 Left: the Netherlands hierarchy of administrative areas, Right the administrative areas in Queensland

There may be many more hierarchical groupings of parcels that are generated by attributes of those parcels: e.g. census districts and areas, postcode, state, country etc. Although The Netherlands and Queensland both use encoding for these hierarchies, they could alternatively be defined as geographical regions.

Geographical region defining a secondary interest.

Many regions in the jurisdiction impact on the property interests. For example, drainage basins, declared weed infestation areas, historic inundation regions. These may follow the base spatial unit boundaries, or may not. In some cases, part of the boundary of a region may be associated with cadastral boundaries, while others follow non-cadastral natural features. (For example, a weed infestation area may be bounded by a ridge line and by a road). These regions have traditionally been viewed as 2D objects, but some 3D secondary interests do exist, for example the restrictions implied by the flight path gradients of an airport.

Secondary interest applied to a single parcel

This is probably the most common form of secondary interest, where multiple parties have interests in the same base spatial unit.

Secondary interest applied to a subset of a base spatial unit

An example would be a defined easement in a base spatial unit, where the geographical extents of the easement are surveyed and registered against the spatial unit definition. There are some subtleties with this kind of interest, e.g. if the base spatial unit is subdivided, should the subset also be subdivided? In Queensland this is not the practice. For example in Fig 7, easement E over parcel 1, on subdivision of that parcel, becomes the same easement over parcels 11 and 12.

Non-base spatial units may be defined as 3D objects. Note that some conventional 2D restrictions may be considered as 3D objects defined in a text-based encoding, for example, restrictions "to the depth of" on a base spatial unit.
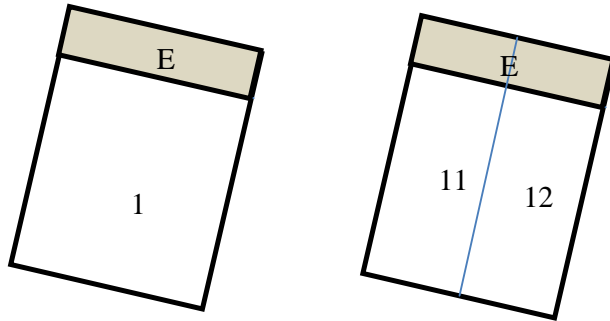
Fig 7 Left - an easement over a single base spatial unit - Right - one easement over the two subdivided spatial units.

Notifications on a base spatial unit

Various secondary interests are recorded which are attached to a single cadastral parcel, but with no detail of the exact location of the interest. For example: 1) An oil pipeline crossing a spatial unit may be shown as affecting a spatial unit. In the Netherlands, a significant proportion of constraints on parcels (more than 90%) (van der Meer 2010) are not directly related to the parcel boundaries. 2) In Queensland, the concept of "road reservation area", where a landowner may have to relinquish a certain area of land for construction of a public road through the property at some unspecified location and at some unspecified future date.

Many legal restrictions that come into force have in the past been "pencilled in" on paper working maps to warn potential interested parties. This concept of "Notings" has been simply moved to a digital storage and discovery mechanism. This type of secondary interest is not without problems, because when spatial are split or merged it is not clear which parts are affected. And after some time the data gets 'polluted' by accidental dropping of interests or allocating them to wrong units (van Oosterom, Maessen et al. 2002)

Network spatial units

In The Netherlands, for pipelines and cables, the legal spaces have become first class cadastral objects of equal level with parcel, apartment rights etc., with their own geometry; raising the important issue of the intersection of 3D spatial units with the base 2D spatial units. (Döner, Thompson et al. 2010). In the Netherlands, a significant proportion of constraints on parcels (more than 90%) (van der Meer 2010) are not directly related to the parcel boundaries.

5. HISTORY IN A DEVELOPING DCDB

It is an important part of any Cadastral database (though sometimes overlooked) to maintain a history of the use of land. In the days of maps on linen, and working maps marked up manually, there was a form of historic record of the cadastre. In moving to a digital cadastre, the Queensland Government, like the Nederlands Kadastre, adopted a form of what is now known as the "Versioned Object" pattern. The LADM also uses this pattern, which permits a permanent but efficient storage of cadastral history within the database itself. As a sideline: the other approach - the apparently simple method of archiving off a copy regularly (say once a year) has proved to be extremely difficult. The archived copy will be in the format of the day, and usually in a structure that requires special software to read it (Sweetkind-Singer,

Larsgaard et al. 2006, Janée, Mathena et al. 2008). McGarva, Morris and Janée recommend to consider "keeping archival data in live access systems"(McGarva, Morris et al. 2009).

In a progressively developing and improving database which records history, it must be recognised that it is expensive and hard to cost-justify the correction of historic information. However, it is commonly accepted that history of "The cadastre as we knew it" is still a valuable resource. Thus historic records of the cadastre may be of a lower standard than current data. For example, in the Queensland case, a view of the DCDB from the 1990's is likely to contain adjacent spatial units with mismatched boundaries. Thus the levels of encoding and accuracy may be non-homogeneous in space and time. As an example, Fig 8 and Fig 9 illustrate the differences between corrections and new data. Note that the historic data is shown in outline only, to allow overlay of the two.

Fig 8 Left: Queensland Cadastre in 2001. Right: Queensland Cadastre now.

As can be seen in Fig 9, there is a combination of additional parcels (A), a change in the definition of the river boundary (B and C), and an adjustment of parcel boundaries (D). The parcels in darker shading on Fig 8 and Fig 9 are strata parcels and volumetric parcels which had not been captured in 2001.

Fig 9 Detail of 2001 superimposed over 2013. Note the change in definition of the river boundary.

6. A POSSIBLE DATA MODEL

The above considerations argue for a data model which can accommodate all levels of encoding, and full 3D parcels. This model should be based on the LADM, and be capable of delivering data in LADM form (perhaps XML encoded), and of accepting such data.

The internal structure of the database may differ in some respects from the LADM model, but this should be kept within control. Some possible areas of difference could be:

1. Additional tables and attributes as per the local profile.
2. Additional linkages and possible redundancies for performance reasons.

6.1. Encoding

The database must clearly be able to accommodate all levels of encoding. It must also have an easily accessible metadata repository which automatically records and displays the level of encoding and the accuracy (and currency) of the cadastre in any area and any era of interest. It must be able to query and display the cadastre at any level of encoding up to the limit set by the data in any area and era of interest. (E.g. If a user wants line-based encoding where the data is polygon based, the conversion should occur automatically)

Fig 10 shows a possible database schema, based on a pilot project attempted in Queensland.

Notes on Fig 10:

1. All relationships are temporal – that is to say, the create/destroy time stamps must be respected in following any relationship links.
2. All identifiers are persistent – that is to say, when a new representation of an object is created in the process of an update, the identifier remains the same.
3. The classes "Boundary", "Boundary3D" and "Corner" are associations, to resolve the many-to-many links in the LADM definition.
4. There is nothing to prevent a LA_SpatialUnit from being bounded by a combination of LA_BoundaryFaceString(s) and LA_BoundaryFace(s).
5. Following the Queensland conventions, a LA_Point is considered to be a 2D point. A Corner is a 3D point along the vertical line defined by the LA_Point. It is identified by an alphabetic suffix, and an elevation. (e.g. corners 2a and 2b are vertically one above the other at the location given by LA_Point 2).
6. RRR details and LA_BAUnit information has not been included in the model at present.
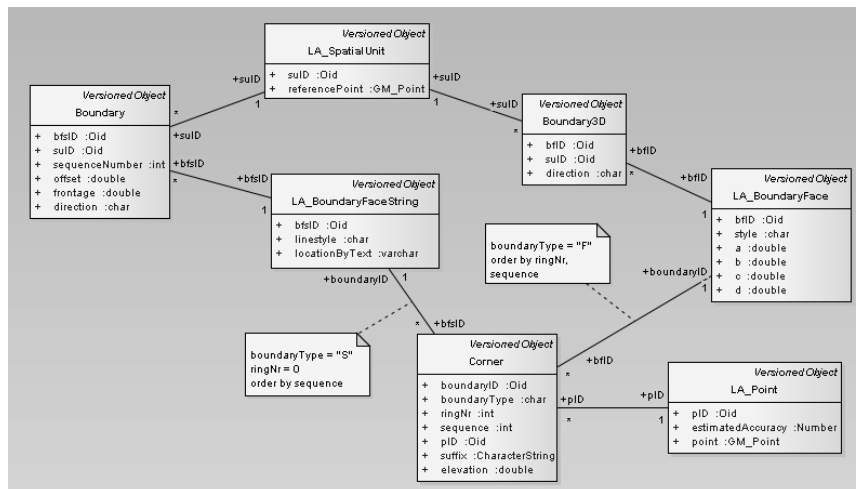7. LA_Point has only been shown in simplified form.

Fig 10 Suggested data model, based on proof of concept database piloted using the Queensland DCDB data.

In the following text, abbreviated forms of the classes will be used where there is no likelihood of confusion – spatial-unit, face-string, face and point.

There is a case to be made for an open-source development of a LADM-based cadastral database, and the software to drive it. This could provide a cheap entry to cadastral databases for a large number of jurisdictions, with a possibility of growing into highly sophisticated databases. The cost of building a DCDB is very high – especially in the later stages, but the principal costs are in the manual correcting and cleaning of the data. This would allow most of those costs to be delayed until some results can be seen from the data captured so far. This software should allow import of data in any of the five encoding levels. There should be the provision to prevent degradation of the data (from a higher level to a lower level), but this should not be a hard and fast rule. If high level data is wrong, it may be better to replace it with correct, low-level encoded data.

Text-Based Encoding

Each spatial unit is connected via boundary records to one or more face-strings which carry the text description. Ideally these would be in anti-clockwise order around the spatial-unit.

Point-Based Encoding

The referencePoint in the spatial-unit records the location of the spatial unit's labeling point.

Line-Based Encoding

The cadastre is represented by a collection of face strings defining the linework, spatial units defining the text labels. Boundary objects are not present. The face-strings are connected via corner records to points. As in the point-based encoding, the labeling points of each spatial unit are indicated by referencePoint.
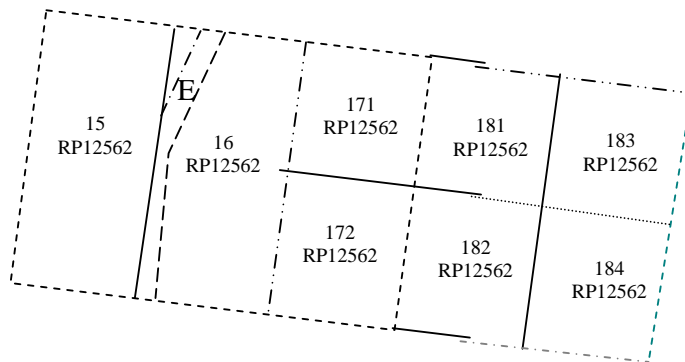

Fig 11 Line based encoding

In Fig 11, different face strings are given different line fonts. There is no connection between the linework and the text. There is also no guarantee that edges do not lie over one another and there may be cases of multiple points at the same physical location. There are well-known parceling routines that can detect and correct gaps, overlaps and mismatches. While it is easy for a person to recognize an easement "E" as a non-base parcel, it may be difficult for a parceling routine to do so, especially in the case of multiple overlapping easements.

Polygon-Based Encoding

Here, each spatial unit connected by boundary records (each of which is marked as being in a forward direction) to the same number of face-string records.
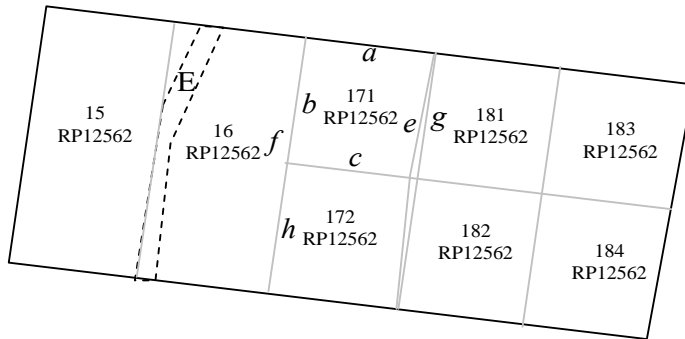

Fig 12 Polygon based encoding.

In Fig 12, for example, lot 171 would be represented as two face-strings, *a*, and *bce*. The separation into two face-strings rather than a single *abce* is to allow for the different attribute (lineStyle) of the road boundary. Note that line segment *f* in boundary of lot 16 does not need to be broken to match segments *b* and *h*. Note also the sliver between *e* and *g*. This is not to say that constraints cannot be built into the database software to detect and even correct these issues, but that the database does not preclude them structurally. In this case an easement is simply a polygon "over the top" of the base parcel polygon.

Note that where there is no requirement for any attribution of linework (no linestyle), the face-strings can be closed to make "face rings" – leading to a truly polygon based encoding.
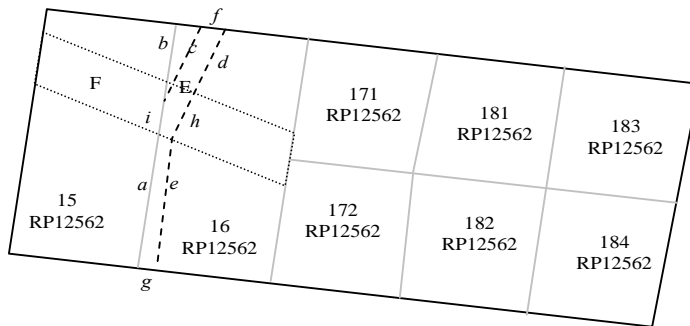
Topology-Based Encoding

Fig 13 Overlaying easements

Here the face-strings are stored once only, and linked to the spatial-unit(s) on the left, and the spatial-unit(s) on the right. In Fig 13, a face-string *ehd* (in that direction) will be linked to easement E via a Boundary record with direction "plus" (or "L" if preferred). *a* will have lot 15 on the left, and lot 16 and easement E on the right. Where easements or other secondary interests overlay, various encoding strategies are possible (see Fig 13): the most parsimonious of these in terms of data storage being:

> Line *i* has lot 15 on left; easement E and lot 16 on right (no linkage to easement F).
> Line *h* has easement E on left; nothing on right (no linkage to lot 16 or easement F).
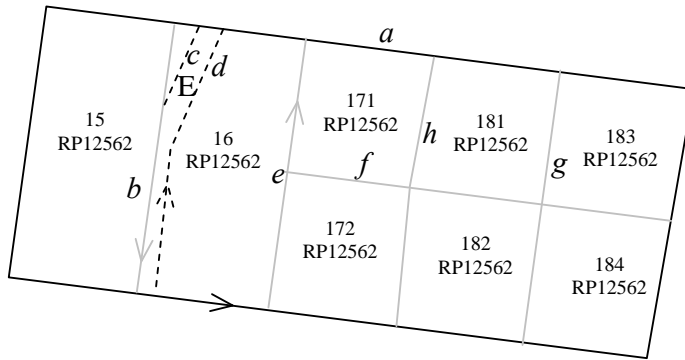
6.2. Generic Encoding

Fig 14 Generic encoding – note that in contrast with earlier figures, the lower case letters are used to label face-strings rather than individual line segments. For example, *a* refers to the outer (road) boundary starting and ending at the top left corner.

This data model also allows a more generic form of encoding which can be mapped from any of the above encodings, and can be mapped to any of them for which the data is of the correct quality. This is in effect, a line-based encoding with additional topological linkages. In this approach, lot 16 would be linked (via boundary records) to face-strings *b*, *a*, *e*, and *a* a second time. Easement E would be linked to face strings *b*, *a*, *d*, *a*, *c*, and lot 171 by *ē*, *f*, *h* and *a*. (where *ē* is face string *e* with a direction of "minus"). The boundary attributes "offset" and "frontage" indicate the part of the face-string that makes up the boundary of the spatial unit.

Table 1 – Allowable conversions of encoding levels

In / Out	Text based	Point based	Line based	Polygon based	Topology based
Text	Y	Y*	Y*	Y*	Y*
Point	N	Y	Y	Y	Y
Line	N	N	Y	Y	Y
Polygon	N	N	V	Y	V
Topology	N	N	V	V	Y

Table 1 indicates the possibility of supplying data at a particular encoding level from a source of different level:

"Y" indicates that the conversion can be readily accomplished,

"N" that it cannot,

"V" means that validation is needed.

"Y*" means that conversion is probably possible, but some research is needed. Clearly, given a topological encoding and a simple spatial unit such as lot 171 in Fig 14, we could generate text such as: "Bounded to the north by Main St., to the east by lot 181, to the south by lot 182, and to the west by lot 16", but it would be interesting to see if useful descriptions could be generated in more complex real-world cases.

6.3. Encoding in 3D

In a progressively developed 3D database, mixtures of encoding will co-exist in the database. This data model can accommodate many combinations. For example:

1. The ground plan of an apartment defined by polygon encoding, and the z-component as a text encoding (e.g. "On Level 5").
2. In Queensland, building unit lots are defined by text encoding (within a base parcel).
3. Horizontal subdivision of a building might be defined by a topological encoding of face strings to define the x/y extents. Horizontal faces may be used to define the z extent, and be defined by elevation only.

In Table 2, various combinations of encoding are shown. Clearly where a fully general 3D object is being represented, the x/y encoding and the z encoding must be the same. However, there are many cases in the cadastral domain where the form of the objects is restricted. A combination of encodings is only acceptable if an automatic conversion to a homogeneous form is possible. (e.g. Y^4 below).

"Y" indicates that the encodings are compatible,

"N" that they are not,

"Y^1" takes the form of a definition of the floor plan of the spatial unit, with a point to indicate the approximate elevation, without specifying the vertical extent.

"Y^2" line encoding in 3D means that space is divided by a combination of face strings and faces, with no attempt to connect them with the spatial unit labelling points.

"Y^3" This is means that 2D space is broken up by face strings, but that certain spatial units have an upper or lower limit imposed on them by horizontal or nearly horizontal faces.

"Y^4" Here, the ground plan is defined as polygons, but the horizontal dividing faces are shared.

Table 2 - Mixed x/y encoding vs. z encoding

2D / 3D	Text based x/y	Point based x/y	Line based x/y	Polygon based x/y	Topology based x/y
Text z	Y	Y	Y	Y	Y
Point z	N	Y	Y^1	Y^1	Y^1
Line z	N	N	Y^2	N	N
Polygon z	N	N	Y^3	Y	Y
Topology z	N	N	Y^3	Y^4	Y

7. SUMMARY

A database built on the model of generic encoding, as described above would require a large array of conversion routines. These could be undertaken in an open-source development project, and while some of them are non-trivial, none require any novel techniques.

The cases chosen in this paper (Queensland and The Netherlands) reflect the background of the author, and a review of this approach with respect to a wider range of jurisdictions would be appropriate. However it is difficult to imagine any cadastral jurisdiction that could not be accommodated by such a model, at least in terms of the geometric layer.

The data model suggested here is relatively simple, but may require additional fields and relations to ensure a fast response. In particular there may need to be some redundancy introduced. Part of the development process would be to try various alternative strategies. An example of the sort of additional data that could be added is a minimum bounding rectangle stored in the spatial unit object, surrounding the points that are in the definition of that unit.

This paper has considered the question of the geographical representation of spatial units within a DCDB, and has indicated that the LADM is a suitable model on which to base the

building of such a database. There is scope to investigate a similar question in relation to the other packages of the LADM. That is to say, do the Administrative and Party packages provide the basis for a general purpose system to support the ownership functions of a land administration package? Can the Surveying and Representation subpackage provide the model for a survey record keeping database?

ACKNOWLEDGMENTS

I would like particularly to thank: Peter van Oosterom for his comments and assistance, and Jaap van Zeevenbergen and Hendrik Ploeger for input on the approaches taken by the Netherlands Kadaster.

REFERENCES

Döner, F., R. Thompson, J. Stoter, C. Lemmen, H. Ploeger, P. Van Oosterom and S. Zlatanova (2010). "4d Cadastres: First analysis of legal, organisational, and technical impact - with a case study on utility networks." Land Use Policy **27**(2010): 1068-1081.

ISO-TC211. (2003, 2001-11-21). "Geographic Information - Spatial Schema." ISO19107, from http://www.iso.org/iso/catalogue_detail.htm?csnumber=26012.

ISO-TC211 (2012). Geographic Information - Land Administration Domain Model (LADM). ISO 19152.

Janée, G., J. Mathena and J. Frew (2008). A Data Model and Architecture for Long-term Preservation. 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL), Pittsburgh, ACM.

Lemmen, C. (2010). The Social Tenure Domain Model: A Pro-Poor Land Tool, INTERNATIONAL FEDERATION OF SURVEYORS (FIG).

McGarva, G., S. Morris and G. Janée (2009). Technology Watch Report: Preserving Geospatial Data. Digital Preservation Coalition.

Sweetkind-Singer, J., M. L. Larsgaard and T. Erwin (2006). "Digital Preservation of Geospatial Data." Library Trends **55**(2): 304-314.

Thompson, R. and P. van Oosterom (2012). Modelling and validation of 3D cadastral objects. Urban and Regional Data Management. S. Zlatanova, H. Ledoux, E. Fendel and M. Rumor. Leiden, Taylor & Francis. **UDMS Annual 2011**.

van der Meer, A. (2010). "Registreer de Wkpb-beperkingen geometrisch: Naar een systematiek die aansluit op de praktijk." Geo-Info **2**.

van Oosterom, P., B. Maessen and W. Quak (2002). "Generic query tool for spatio-temporal data." International Journal of Geographic Information Science **16**(8): 713-748.

APPENDIX I – SMALL PARCELS

In Queensland there are special lots known as "Access Limiting Lots", which are intended to prevent the construction of an access from a private property to a road (for example to a freeway). There are typically 100mm wide (ground scale). Although not now being created, there are several thousand of these still current. More recently, small details have been

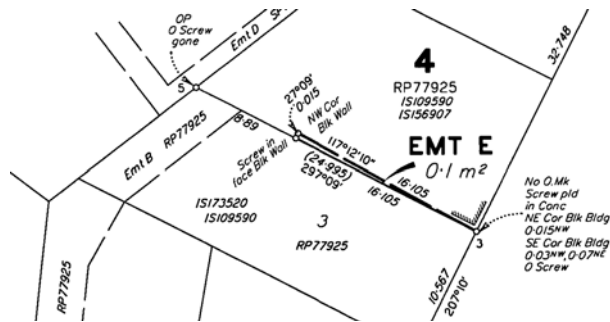included in plans – including the definition of a series of extremely narrow triangular easements – see Fig 15.

Fig 15 Detail from a survey plan showing an easement which is 15mm wide at its widest point

These set a limit on the "snapping distance" that can be used in any automatic correction scheme, but a snapping distance of 2mm at ground scale (as used in the Queensland DCDB) allows small errors such as shown in Fig 1, that would be invisible at a typical map scale (e.g. 1:2500), not to be corrected.

BIOGRAPHICAL NOTES

Rod Thompson joined the Queensland Government in the Department of Mapping and Surveying in 1985, and was instrumental in the development of the Digital Cadastral Data Base (DCDB). He gained his doctorate in 2007 with a dissertation entitled "Towards a Rigorous Logic for Spatial Data Representation". He remains employed by the same (albeit re-named) department, but also retains a research connection with the Delft University of Technology.

CONTACTS

Rod Thompson
Delft University of Technology and
Department of Natural Resources and Mines
Landcentre
Main and Vulture Streets
Woolloongabba Q 4102
Australia
Tel: +617 38963286
E-mail rod.thompson@qld.gov.au
Website: http://www.dnrm.qld.gov.au