The Development of GEOSCAPE: A National Database of Building Outlines

Brian Burbidge
PSMA Australia
Brian.Burbidge@psma.com.au

ABSTRACT

This paper focuses on the development of a new product called GEOSCAPE, which is a national representation of building outlines together with additional building information. The paper looks at the development of the product from concept through to the reality of a consistent national product, integrating contributions from a wide range of data sources to produce a product that has an extensive range of uses. Projects like these are more than a technical challenge, so the paper also outlines the process of developing the product along with the depth and breadth of skills and disciplines required. It also presents some of the interesting spatial challenges and learnings gained along the way that may be of interest to professions working in the spatial industry as well as users of these products.

KEYWORDS: *Spatial data, database, product development.*

1 INTRODUCTION

PSMA Australia offers a national asset of quality spatial information derived from authoritative data sources. These national location datasets underpin an ever-expanding range of business solutions and government services. Our flagship product, G-NAF, is Australia's only authoritative, geocoded physical address file. G-NAF is complemented by datasets of roads, cadastre parcels, administrative boundaries, features of interest and more. Used together, or separately, PSMA datasets provide the geographic context that enables effective decision-making and innovation based on quality location data.

PSMA Australia and its clients have been exploring the possibility of expanding its product range to include a national representation of building outlines to supplement the existing products. This product is to provide a new nationally consistent window into the built environment and a mechanism for more readily integrating the existing products and other information through different lenses.

Opportunities to develop products from a blank canvas are few and far between, so the strategy and approach adopted needed to be flexible and iterative. The development of requirements is constantly being tested against the hard reality of what is achievable in a reasonable timeframe and budget.

Over the last two years, PSMA has been developing the concept and exploring ways the product can be developed over Australia either through our existing spatial data channels or through other sources available in Australia. In parallel, a similar exercise has been underway identifying the key applications and client requirements as well as exploring the technologies and approaches which could be used to develop the product. PSMA was also aware that others here in Australia and overseas were looking at developing similar products and services although from different lenses and perspectives about the built environment.

2 GEOSCAPE THE PRODUCT

From the outset it was understood that GEOSCAPE must integrate with existing products developed by PSMA and help provide an interface into the built environment. Roof and Oleru (2008) defined the built environment as the human-made space in which people live, work and recreate on a day-to-day basis, encompassing places and spaces created or modified by people including buildings, parks and transportation systems. The existing PSMA products and services currently reflect parts of characteristics articulated and expected by users interested in the built environment (Table 1). The structure of GEOSCAPE is simple and incorporates a number of components that can be expanded into the future (Table 2).

Table 1: Existing PSMA products and their characteristics.

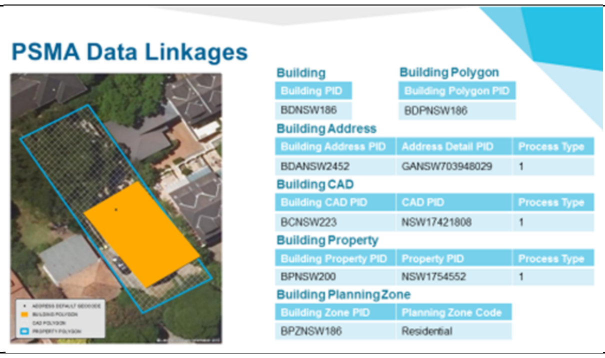
Product	Characteristic
CadLite	More than 14 million cadastral polygons representing over 10 million cadastral and property features.
G-NAF	More than 13 million geocoded addresses generated from over 30 million addresses from 10 contributors.
Transport and Topography	More than 2.7 million km of road centreline plus railways, airports, greenspace and hydro.
Features of Interest	More than 8 million point and polygon features.
PSMA Cloud Service	Daily refreshed databases providing address verification and spatial capabilities across key government and commercial services.

Table 2: Selected GEOSCAPE components.

Building outlines: • 2D roof polygon. • Roof area. • Roof pitch/complexity. • Ground level coordinates for roof vertices. • Number of roof vertices. • Ground level building centroid. • Maximum roof height. • Swimming pool indicator. • Roof material. • Solar panel indicator. • Residential land use indicator.	
Landcover: • Impervious surfaces: o Built up areas. o Road and path. o Bare earth. o Buildings. • Vegetation: o Tree coverage. o Grass coverage. • Water.	

Linkages:

- G-NAF.
- Cadastre.
- Property.
- Zoning.



2.1 GEOSCAPE Indicative Uses

Initial concepts envisaged GEOSCAPE being used across a range of industries and sectors including utilities, finance, agriculture and emergency services. The uses that have been identified to date far exceed these initial expectations and the list is not complete. The visualisation of the GEOSCAPE concepts has not been fully explored, however a couple of concepts were developed from the 2.5D products which we are sure will be developed more professionally by the PSMA VAR network or direct clients. Figure 1 shows some very basic outputs developed from the first pilot study over the Chatswood region in Sydney. They give an indication of what can be produced from the data. This incorporates building outlines displayed as 3D features using feature attributes along with shading based on roof complexity and roof type characteristics in addition to 3D trees/vegetation representations.

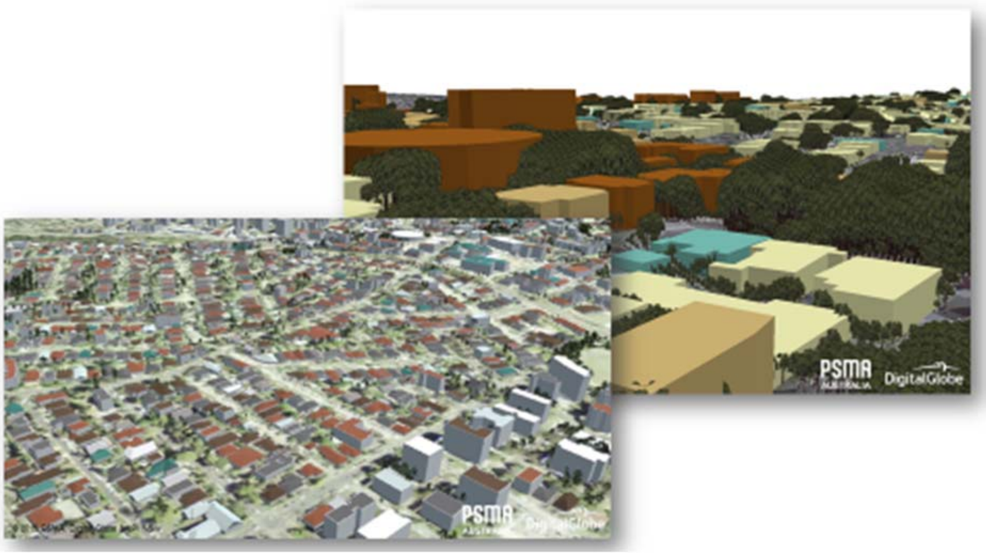


Figure 1: Examples of GEOSCAPE products.

Based on what we can see, the benefits of the data go beyond the simplistic 3D representations as illustrated in these images. We understand from some clients the data is likely to generate operational efficiencies without further modelling, with many further benefits that could be realised through modelling.

3 THE DEVELOPMENT APPROACH

The development of the product has been very iterative, involving both business partners and technology providers in an attempt to develop the most flexible product:

1. *Product concept* – initial development of the product concepts were explored through high-level consultation with the potential end users, data custodians, a range of experts, industry colleagues and the PSMA Board. From the outset the project looked technically achievable, however gauging the maturity of the market and balancing expectations against realistic targets, budgets and timeframes was the initial challenge.
2. *Desktop research* – was undertaken to review research and pilot approaches from within Australia and around the world. This also included further consultations with subject experts to provide advice and guide on feasibility around production techniques and strategies. This drove home the additional challenge that such a project has not previously been undertaken nationally at this level of detail. It also identified the diversity of data that exists within Australia at the government and private company levels. Again, the challenge has been around understanding the data and its suitability for inclusion in this type of product and project.
3. *Pilot projects* – Three pilot sites (about 16 km^2 in size each) in Sydney and Brisbane were chosen to test the technologies and prototype product models in the Australian context. From a technology perspective, the pilots tested satellite technologies marketed by Digital Globe and Airbus, as well as aerial approaches utilising LiDAR and traditional photogrammetric techniques provided by local companies. The pilot areas were chosen to test a range of complex structures (residential, commercial, industrial and high-rise buildings), terrains and vegetation forms.
4. *Product and production specifications* – Following further market testing of the pilot products, the full national production and testing specifications were developed for both the 'urban' and 'rural balance' areas. This was the most resource-intensive phase involving both the preferred contractor (Digital Globe) and key client groups. As with all PSMA production processes, documenting and understanding process flow is critical, particularly given the production can be done across different technology platforms, companies and a need to fit in with other PSMA product delivery schedules.
5. *First production area* – As the first stage in the production implementation, an area of approximately 16,000 km^2 surrounding Adelaide has been initiated. This will test the full production methodology and provide an opportunity for a final review of the specifications. This will be completed by early April 2016.
6. *Full production* – The full national coverage is planned to take a further 15 months from the completion of the Adelaide site.
7. *Maintenance* – A rolling maintenance schedule will be implemented to coincide with PSMA's existing quarterly maintenance schedules as well as specific client requirements.

4 PROJECT COMPLEXITIES

There are a number of issues and interesting points that arise from a project like GEOSCAPE. Some are not new to PSMA but take on a different perspective with this product. Some, which are more relevant to the surveying and spatial information community, are outlined in this section.

4.1 Diversity of Skills

Diversity of expertise and skills is essential in scoping such a project, and certainly the consultations and planning phases to date have involved the usual GIS, surveying and associated technical skills, but also a healthy mix of IT, databasing, business, project management, financial and crowd-sourcing skills.

The area where this has been more challenging has been around generating a common base level of understanding about the product across all the participants who carry their own specialised lenses into the product and its production. This has been made more complex as the product concepts have evolved and a deeper understanding is gained about each of the components required. This is hard enough when everyone is in the same organisation but certainly more complex when you are dealing with different experts and organisations around the globe. Regular structured communications through teleconferences, webinars and internet-based project and document management tools have contributed a lot in managing everyone's understanding.

4.2 Product Categorisation of Australia

One of the business drivers has been to develop the product based on some of the priorities as identified by the key clients. These priorities generally reflected population densities but this was not necessarily uniform across Australia. A critical part of the product research identified there were different product characteristics across Australia. However, the method of classifying these differences was not consistent and a uniform approach could not be easily developed. Some of the characteristics required varied with topography, population density, standard administrative boundaries as well as different proximity criteria.

An initial assessment of existing data sources indicated there is a general lack of reliable information nationally about infrastructure and the built environment. Obviously, for this type of project the location and density of infrastructure is critical along with an understanding of the business drivers for end user applications. An initial assessment and categorisation of areas was undertaken using a combination of population based on the ABS Australian population grid (ABS, 2014) and an understanding of some risks that could be related to other geometries. This was sufficient to broadly classify Australia and to determine the approximate size of each category and distribution around Australia. Figure 2 illustrates that over 90% of Australia's population lives in less than 1% of the Australian landmass, which also largely reflects the majority of the infrastructure that is being targeted in this product.

Statistics based on the ABS data suggest that of the approximately 30,000 km^2 where 90% of the population exist, one third of the areas are communities greater than 1 million, the middle third cover populations from 4,000 to 1 million people, and the remaining third covers populations less than 4,000 people. Populations over 1,000 are classified as 'urban' under this initial capture program developed by PSMA.

While it is important to understand where 90% of the population are located, it is a slightly more complex approach to identify where the remaining infrastructure is located. The ABS population grid focuses on people and residential characteristics of that population. It is less reliable when this is extrapolated to identifying the existence of infrastructure. The population grid does not reflect the location of low population and rural infrastructure well. Likewise, the population grid is based on the 2011 census and in some areas (primarily fringe urban) there has been substantial change in the last 5 years.

While PSMA initially used a number of criteria to develop five categories over Australia, this has been condensed to two categories being 'urban' and 'rural balance'. This change evolved as a greater understanding was gained about specific use cases and the accuracy levels expected by the users in each of the categories. Simplistically, within urban areas there is an expectation that the location of building and the depiction of building will be higher than those in more rural locations. This is an entirely reasonable expectation given the linkages that will be developed to other PSMA products.

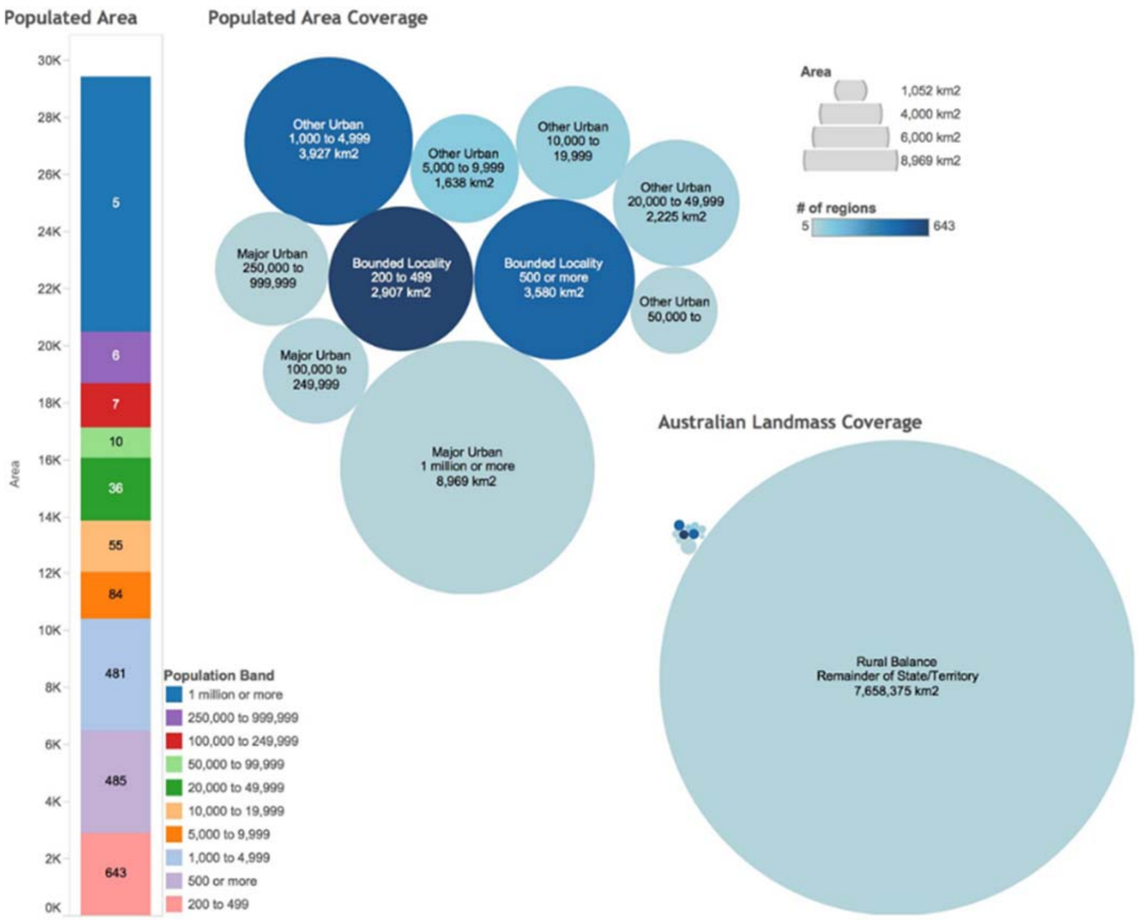


Figure 2: Analysis of population density across Australia, derived from the ABS Population Grid 2011.

Based on a greater level of understanding about the end user requirements, it has been necessary to undertake a further review of the categorisation of the country to ensure that significant populations and associated infrastructure has been classified within the urban areas (Figure 3). The analysis has been undertaken with limited additional reliable information other than data PSMA has access to from other sources. The process used was neither scientific nor repeatable and included both visual interpretation of the most current imagery available as well as other derived data where we could derive associations from our other contributions. It is estimated the second iteration may have increased the percentage of structures in the urban classification by approximately 1.5% over the previous attempt, and it has generally created more uniform coverage over the main metropolitan areas and condensed the area covered around individual smaller communities (generally by removing areas where infrastructure was not present). Figure 4 illustrates the urban classification over Sydney.

In total, it is estimated there are in the order of 25 million buildings represented nationally in this product and that the maintenance schemes used can be better targeted once we understand the current infrastructure. While it is generally understood that Australian urban areas are growing, we do not necessarily understand where the growth areas are outside the main metropolitan areas. Certainly the mining boom has generated a lot of growth in Western Australia and Queensland over the last decade. In some cases, in these states, complete communities have been constructed in less than 5 years.

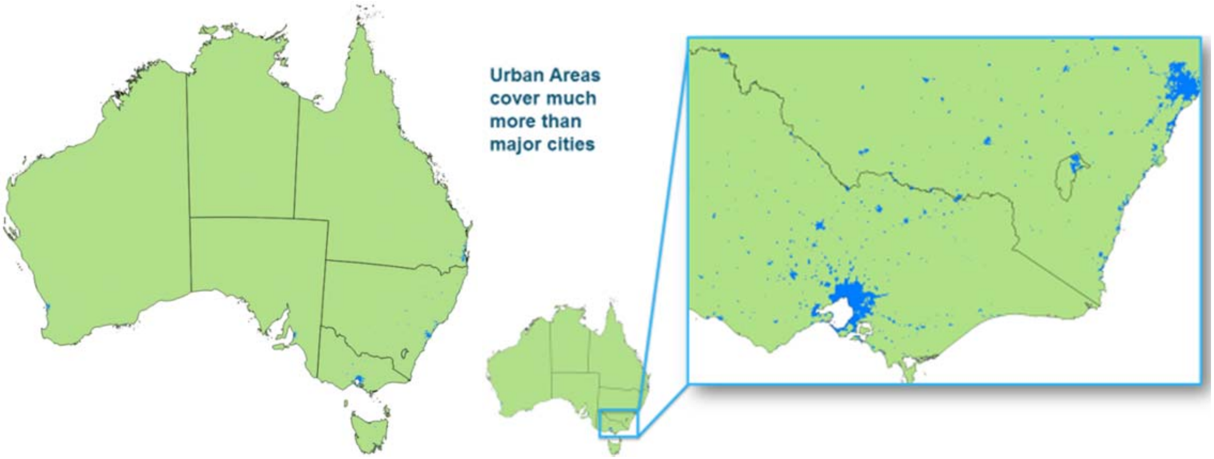


Figure 3: Distribution of urban centres in Australia.

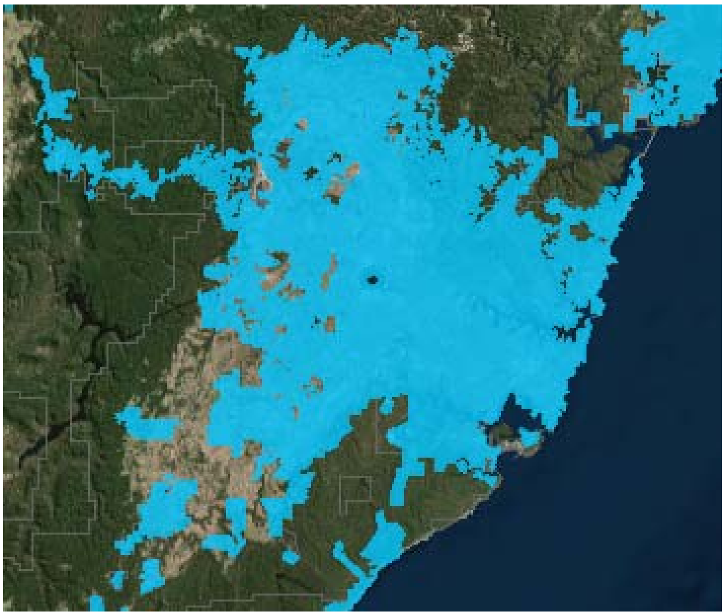


Figure 4: Urban classification across Sydney.

4.3 Quality of Building Representations

As part of the preliminary investigations, the difference between satellite imagery, LiDAR and traditional photogrammetric extraction techniques using 3- and 4-band imagery were tested to determine the differences and qualities of each approach. Differences were identified in the resulting representations, but they were not considered significant. Hence the decision was made use high-resolution satellite imagery provided by Digital Globe. This solution is likely to provide greater opportunity to develop a nationally consistent product, which is not possible with the other approaches tested. Figure 4 demonstrates the satellite solution utilising high-resolution imagery over the urban areas (yellow shaded areas) and mid-range imagery for the rural balance (unshaded areas).

While the product will initially only utilise satellite data, there is no reason to limit the range of alternative input sources into the future. Certainly the initial review of the outputs from LiDAR (captured at 1 point per m^2) suggests this is not dense enough to provide consistent building outlines. Unfortunately, the majority of LiDAR captured across Australia at present is at this lower end of the point density spectrum. It will be interesting to look at the outputs from the ACT LiDAR capture program to see what can be achieved with a much higher point density.

Likewise, the techniques used to develop products from aerial photography showed considerable variation across study areas, possibly due to a combination of factors which may have proven difficult to control in a large project. Certainly to use these techniques would require extensive use of existing data archives as well as new imagery where each project would carry different characteristics and qualities reflecting its original purposes. This carries the risk of a higher range of output qualities, which may impact the application of the end products. This can also be an advantage, however in the initial stages of developing a national product it was seen as a large challenge to manage nationally.

The local companies who assisted with this assessment certainly demonstrated a high level of expertise, knowledge and honesty. It is hoped that PSMA can find ways to utilise their services into the future.

4.4 Product Maintenance

Whilst the initial data capture is a relatively straight forward process, the process of product maintenance is not and presents a number of challenges. There are two key variables that need to be captured and managed:

1. Urban expansion, which is likely to occur on the fringes of the existing urban areas, and a combination of change as a result of urban infill and renewal.
2. Changes in land use, which could result in changes to infrastructure in both urban and rural areas.

Some of the informal testing of the level and location of change suggests that some can be more easily predicted while others are generally more difficult to determine. In the latter cases, they may be near-spontaneous or may not materialise for several years. PSMA already has to deal with these apparent contradictions in the existing products, so GEOSCAPE is likely to assist users in rationalising some of the differences.

The refresh frequency required to map the changes will largely be reflected by the end users. However, there is a commitment to maintain the data on different frequencies, ranging from annually to once every 4 years. Again, this will depend on users, uses and geographic location. No one technique will identify all forms of change and the subsequent need for maintenance satisfying all users. PSMA is looking at a range of techniques utilising grids to synthesise areas of change and user requirements to help form the maintenance program. Some examples of different approaches are illustrated in Figure 5.

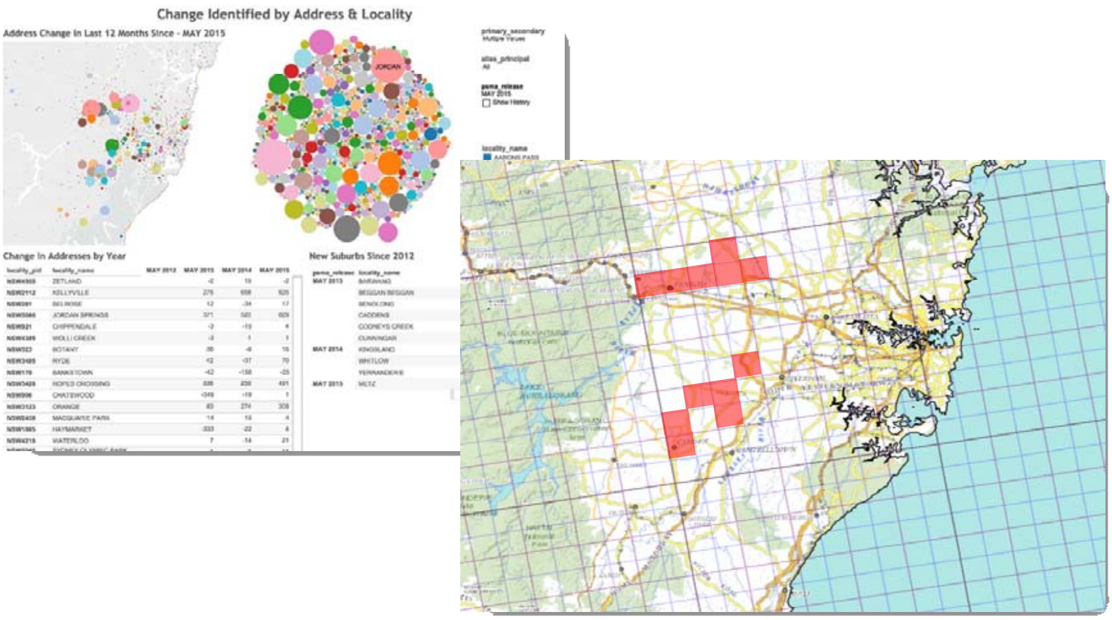


Figure 5: Change detection by locality and maintenance by geographic grid.

4.5 The Cadastral Challenge

One of the challenges facing PSMA is how to rationalise the differences within and between the cadastres across Australia. We know there can be significant spatial inaccuracies in different locations across all the cadastres. In the past, most users have understood they exist, and users have developed work-arounds to overcome the differences. Most GIS operators have seen the differences between road networks and the cadastre, but these differences can often be ignored in applications. In GEOSCAPE, these spatial differences will be evident as there will be a consistent spatial accuracy across Australia that will not be present in the cadastres.

The other challenges with each of the cadastres and property representations will be differentiating ownership from other rights, and accommodating vertical interests. In terms of ownership and rights, most of the cadastres treat these differently, and there is a general assumption that GEOSCAPE will assist is identifying relationships between ownership and some rights to the physical structures present on a site. GEOSCAPE will show some relationships between property and buildings.

In some States and Territories there is very good differentiation between vertical interests, such as differentiating various ownerships in a high-rise development. Unfortunately not all cadastres have this level of sophistication. In other situations the cadastre may only represent one feature being the ground level base parcel with little or no reference to vertical developments and ownerships. Some of these differences have already been addressed in PSMA's existing products whose integration with building outlines appears to be consistent and legitimate from existing testing.

The differences also extend under the ground with some cadastres representing interests such as carbon sequestration interests. These are very legitimate uses for a cadastre, and PSMA fully supports their inclusion as long as they can be easily identified and differentiated from other forms of ownership, rights or interests.

Other challenges arise where rights are used to handle developments such as large infrastructure projects either owned/licensed privately or controlled by the State/Territory, e.g. railway traversing across other forms of right or ownership. In some cases, these may not be represented in the cadastres or may not be easily differentiated within the cadastre. Some of these may be historical artefacts such as multiple occupations on a single Crown entity. In some cases, these Crown lands units/entities may not be titled or treated as a rateable entity and may or may not be addressed. In the Sydney pilots, the largest number of structures that could not be associated with an address were structures controlled by local councils (sporting and other facilities), churches, and complex facilities such as hospitals and universities.

The GEOSCAPE pilot studies have also helped to identify a very small number of inconsistencies in the cadastres and associated addresses. In these particular cases, these have been fed back into the respective State and Local Government agencies for clarification and or correction. These are all opportunities and challenges which are likely to be encountered through the development of GEOSCAPE.

4.6 Utilising the Crowd

One technique which has not been previously used by PSMA is crowd sourcing. GEOSCAPE will utilise crowd sourced information in the production processes. There are different ways to utilise the ‘crowd’, and in recent years crowd sourcing has proven very successful in providing quick responses to humanitarian disasters around the world. A lot has been learnt from these campaigns in both what can be collected and how to present information to the crowd to maximise the chance of them being able to provide a valid judgement. In many ways crowd sourcing is similar to consumer market researching, where it is important to ask the right questions in the right setting to solicit measurable and meaningful responses.

The pilot studies utilised crowd sourcing for solar panels and swimming pools, and the results were promising. Crowd sourcing works well where there is a visual interpretation required, and the use of imagery can be very powerful for a lot of applications. Given the size, shape and form of both swimming pools and solar installations, crowd sourcing was seen to be well placed to identify and differentiate these types of structures.

Depending on the type of interpretation required, there may be different characteristics required to be present in the imagery. In the case of swimming pools and solar panels, it is possibly best to use mid-summer imagery where there is little impact from shadowing of the feature. For the pilot area, winter imagery was used, which unfortunately meant pools could be empty, obscured by pool covers or could be shadowed by nearby vegetation. This made it more difficult for the crowd to differentiate between features in the imagery. The crowd was also presented with representations from more than one dataset. In a few instances where there were slight misalignments between the representations, the crowd was not able to consistently determine/rationalise what they were being asked to assess.

This all illustrated the need to be very careful about what information is used in crowd sourcing and that structuring a campaign well requires a well thought-out plan taking into account the perspective of the crowd. The benefits of crowd sourcing are that the crowd is multicultural, global, works 24 hours a day, is passionate and very good at this type of activity. While the pilot area was limited in size, the information collected from the crowd was very useful and has provided valuable learning for the real production process.

5 PRODUCTION TIMEFRAMES

Originally, this was seen as a project that could be completed relatively quickly and the business case would only take 4 months. In reality, the business case has taken over 12 months to prepare in parallel with other activities. Currently, PSMA is working with Digital Globe on the production of the next milestone being the development of the Adelaide Production Area. At this point, the first production stage has commenced over the Adelaide region and this should be completed in April 2016.

Following the Adelaide region, the next milestones are:

1. Two more major capital cities (urban areas) by June 2016.
2. The remaining capital city urban areas by September 2016.
3. The remaining urban areas (nationally) by March 2017.
4. The rural balance by June 2017.
5. Maintenance program commences in September 2017, using target grid cells and an annual national land cover refresh for all urban areas.

6 NEXT STEP: ADELAIDE

The Adelaide production area extends north to Port Wakefield, east to near the River Murray and south to just east of the River Murray mouth. This area is seen as a test of the entire production process as it will include both the main metropolitan area as well as the surrounding smaller urban centres. In total, this production area is some 16,000 km^2. Figure 6 illustrates the Adelaide production area (red) urban classified areas (blue) and the Digital Globe World View 3 image footprints (blue). The image capture was not fully completed in this image.

Figure 6: The Adelaide Production Area.

CONCLUDING REMARKS

GEOSCAPE has been an opportunity for PSMA to plan a product from a blank canvas using the knowledge that has been accumulated over the past decade from its other national products. It is utilising new sources of data and developing production techniques that will help integrate and complement the existing product suite. GEOSCAPE is one lens into the built environment with very specific initial objectives, however it has been developed in such a way as to maximise its flexibility and to enable the product to evolve and provide a scalable platform into the future.

The development methodology has been rigorous, iterative and involved all stakeholders and clients throughout the initial phases. This has extended the time to develop the product prototypes and full specification and production methodologies, however the full production phases and future developments should all benefit from the rigorous approach adopted.

While Australia has an urbanised population, the built environment is not as urbanised and is not currently well understood or represented nationally. This product is designed to provide a more nationally consistent representation of building outlines to help raise the understanding of this part of the built environment. Initial indications are that it will achieve its objectives and provide a wider range of benefits to industry and the wider community than initially anticipated.

REFERENCES

ABS (2014) Australian population grid 2011, http://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.007 (accessed Feb 2016).

Roof K. and Oleru N. (2008) Public health: Seattle and King County's push for the built environment, *Journal of Environmental Health*, 71(1), 24–27.