Delivering Sustainable Data and Product Management Business Processes

Lesley ARNOLD, Landgate, Australia

Key words: spatial information updating, geographic data modelling, cartographic processing, generalisation, data management, map publishing, sustainability

SUMMARY

The traditional map production process is resource intensive and time consuming, and the information collected is often trapped in formats that prevent it from being used in other products and applications. Work flows are often divided and therefore difficult to manage, and data currency is inconsistent across an agency's data sets and products.

To address these issues, Landgate is developing sustainable data and product management business processes which allow digital data products to be updated in real time. This is a significant reform of current mapping processes. The data and product management framework incorporates timely data collection techniques, a multipurpose geographic knowledge-base, automated mapping techniques and Dynamic Spatial Updating (DSU) methods.

The approach will provide the flexibility to deliver new products (and spatial updates) to the user community more quickly. The overarching strategy is to capture geographic features and real world concepts once only; and use this geographical knowledge to create multiple map products in a variety of formats, themes and scales to meet end user needs. Sustainability is achieved by integrating workflows, automating existing tasks and encouraging customer participation at each stage of the product lifecycle.

The Dynamic Spatial Updating (DSU) process is central to delivering the automated components. DSU refers to the modelling techniques, spatial queries and algorithms that are applied to automate map product revision. DSU is implemented in a database-sharing and spatial (map) views environment. This is a paradigm shift away from managing isolated product data sets. Spatial views are not separate digital files (in the traditional sense) but rather live windows to the underlying geographic source data. This means that updates performed in the database are immediately available (and transparently) to a user's spatial view.

Delivering Sustainable Data and Product Management Business Processes

Lesley ARNOLD, Landgate, Australia

1. INTRODUCTION

In the last decade, national mapping agencies have moved to geographic information systems (GIS) to take advantage of the new tools for improved data management and spatial analysis.

Whilst this has been a huge technological advancement for data custodians, analysts and decision makers - map publishers and product users are still faced with lengthy production times and out-of-date data and maps in the market place. It can take 6 months for a change in our environment to be effected in road maps and navigation systems.

Today's information users are demanding immediacy with respect to data currency. Given current technological advancements and access to the latest communication devices, this is not perceived by the user community as unreasonable. However, for data custodians and managers of spatial information the process involved in reflecting a geographic change in databases and products, the instant the change occurs in the real world, is complex.

The challenge ahead is to develop a framework that will enable a vision where geographical changes in our environment can be reflected in databases and downstream products in real time. Landgate has taken up this challenge and has developed a data and product management model where spatial information can be streamlined and automated. The framework considers the life cycle of geographic features as they change overtime, and how they are managed in a digital environment and translated to products for public consumption.

A key criterion in the development of the framework was to consider the long term sustainability of the methodologies. The maintenance of geographic data is resource intensive and with existing processes, data currency can only be improved with the allocation of more resources. As a consequence, the new approach had to embrace task automation, avenues for data collection and maintenance partnerships, data structures that evolve with changing user requirements over time, and an IT environment that allows for economical renewal and supports reuse of business workflows, data and information processing.

This paper explores the components of the Sustainable Data and Product Management Framework (SDPM) from the creation and evolution of a geographic object in the real world, to its various portrayals in digital and hard copy map products.

2. BACKGROUND

Landgate is transitioning its cartographic production from a desktop publishing (DTP) environment to GIS. The cartographically enhanced data will not be stored as separate digital data files. Instead, cartographic products will be derived directly from the source data held in the GIS environment. This means that products will be updated the instant an edit is made to the primary data source. This constitutes significant business reform from existing traditional methods where maps are created as individual data sets and require updating in parallel to the source data.

In 1993, Landgate (similarly to other mapping agencies) adopted a combination of desktop publishing (DTP) and computer aided drawing (CAD) packages for graphic presentation to achieve economies in producing traditional paper map products. At that time, early GIS and Spatial Data Management Systems (SDMS) lacked sufficient and flexible design capabilities and fine graphics control for achieving output quality (Ertle and Lauinger, 1995; Beard, 1993; Dobson, 1988; Andersen, 1997). GIS also had incompatibility issues with high-end imagesetters utilised by the printing industry (Leskinen, 1995; Sena, 1991; Peck, 1995).

Whilst DTP and CAD systems proved to be a vast improvement on creating thematic map products by hand the spatial information updating process became far more complex. This is because data is stored across a large number of individual files and software environments making it logistically difficult to manage.

Multiple data sets are required (one for each scale range and map product) to manage both the true geometric and the cartographic representations of the same data. Whilst this has provided versatility as data provider and map publisher, the overhead of managing multiple data sets is resource intensive and not sustainable in the long term (Figure 1).

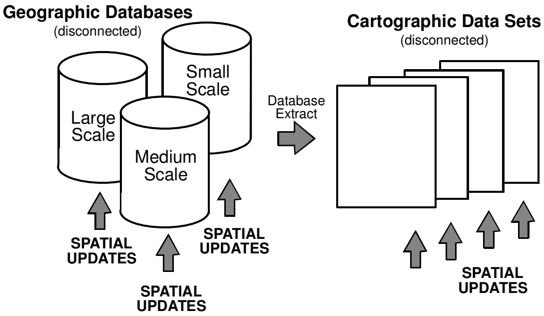

Figure 1: Traditional map production and revision processes (Arnold, 2006).

The first generation DTP and CAD systems mimicked the conventional map production process and, as a consequence, there is little attribute data other than information required for symbology. These product data sets, whilst having a wealth of graphical information, cannot be used for navigation, routing and location-based technologies. In addition, data is not wholly aligned with contemporary data formats and new computing paradigms. This restricts usability in an online environment.

With all the problems inherent in DTP and CAD many agencies continue to use these systems. This is primarily because the conversion of data and products from one computing paradigm to another is achieved at considerable cost and, therefore, a significant return on investment is warranted. In moving to GIS, the methodology had to be able to offer considerable productivity improvement and the ability to create contemporary and innovative products - and not be just another way to produce maps.

The full potential of GIS lies in the ability to use analytical (not just graphical tools) to derive maps, and the ability to automate the map production and revision processes using GIS capabilities. These tools along with other data maintenance strategies have been encapsulated in Landgate's Sustainable Data and Product Management Framework in order to manage the lifecycle of geographic features in a more sustainable way.

3. SUSTAINABLE DATA AND PRODUCT MANAGEMENT FRAMEWORK

The Sustainable Data and Product Management (SDPM) Framework is essentially a plan for managing geographic data resources and products more efficiently to meet growing community needs for accurate and current data. The concepts and design underpinning the framework are driven by the need to:

- provide an integrated workflow from the capture of a geographic feature to its various representations in multiple products;
- provide accurate, current and complete geographic data coverage across the State of WA;
- create an extensible multipurpose GIS for managing geographic information complexity;
- enable the flexibility and scalability to create a wide range of products from the primary data;
- automate the map production and revision cycles so that more products can be produced in less time;
- enable customisation of thematic map products by users in an online environment;
- deliver readily accessible mapping products in a wide range of formats to support current and future community needs; and
- focus on customer participation in all stages of the data and product management lifecycle to ensure data and products are fit for purpose.

3.1 Sustainability Principles

The SDPM framework incorporates the principles of sustainability. This is essential. The current geographic data management and map production environments are not sustainable in the long term. The present use of human and computing resources can not be continued into the future and the ability to create products and services for future generations is curtailed by inappropriate business processes and out-of-date technology. In addition, the technical skills of staff are progressively becoming outdated and new recruits difficult to resource because of dissimilar skill sets.

In focusing on sustainability principles the framework will allow for:

- the development of real time business processing to ensure the provision of comprehensive and fully maintained geographic information for effective social, economic and environmental decision making;
- the use of robust, energy efficient and cost effective systems that enable contemporary data validation, economical technology renewal; and reuse of data, information and workflows;
- task automation, to address human and technology resource constraints; and
- customer participation, to ensure the longevity and usage of the agency's data resources and products well into the future.

3.2 Focus Areas

The SDPM Framework has five key Focus Areas. They are Data Compilation and Capture, Data Management, Map Production and Revision, Product Development and Customer Delivery. These relate to stages in the data and product management lifecycle – from data capture through to the provision of products to customers (Figure 2).

Figure 2: Focus areas reflect stages in the data and product management lifecycle.

A key driver for achieving sustainable objectives is to apply automated techniques to not only develop products more efficiently, but more importantly, to ensure that once products are developed they can be continually updated in real-time. To achieve this, workflows will be fully integrated across the key focus areas (Figure 3).

- Timely Data Capture: Application of effective methods for collecting the changes that occur in our environment. Supply-chain business modelling will advance data collection techniques from traditional manual solutions to more integrated and automated workflows.
- Integrated Systems and Data: An environment of integrated databases where geographic features are interrelated at the feature level. This supports value-adding opportunities and enables data analysis, and search and querying. It includes linking a feature to its name and address, and potentially other data types, such as plans, land valuations, cadastral parcels, indexes, pictures, audio and 3-D visualisations.
- Map Production and Automated Revision: Task automation includes the application of map labels, symbols and marginalia; and the implementation of dynamic spatial updating techniques that allow database updates to be transmitted automatically according to specific product designs.
- Flexible Product Development: Spatial updates to map products are triggered automatically to various digital product forms and online delivery environments. Hardcopy products can be downloaded and printed from updated map data at anytime – meaning that hardcopy products are always current at the time of printing.
- Effective Customer Delivery: Spatial updates are triggered automatically to standard and user-customised product derivatives.
- Customer feedback and collaboration: This customer participatory approach allows customers to drive data requirements and priorities. It fosters customer collaboration opportunities and enables feedback to be effectively incorporated throughout the data and product management lifecycle.
- Real-time Product Updating and Delivery: While data capture will be timelier, once an update is entered into the primary source databases, modifications to geographic features will be delivered to products in real-time.

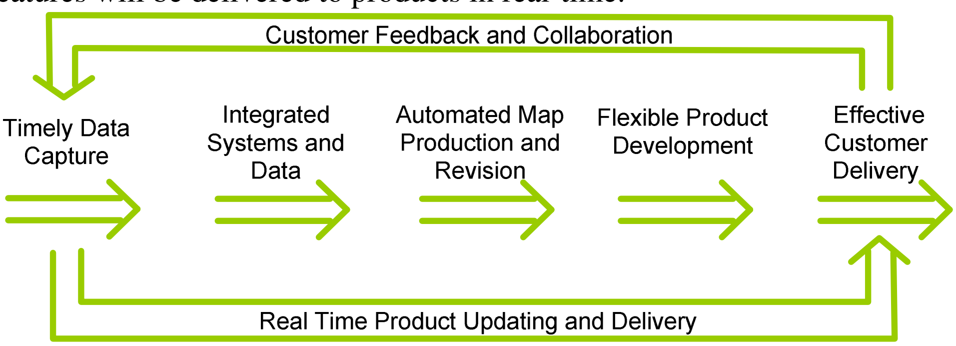

Figure 3: The SDPM Framework consists of fully integrated workflows (Landgate, 2009).

3.3 Strategic Delivery System

The Strategic Delivery System (Figure 4) represents a process view of the SDPM Framework and encapsulates the concepts of integration, automation and customer participation. The model will enable Landgate to move from a data maintenance cycle based on hardcopy product lifecycles to continuous real-time product updating. It will also allow Landgate to move from standard map products and limited data customisation capabilities, to products based on customer driven data and presentation requirements and maintenance priorities.

The Strategic Delivery System illustrated in Figure 4 highlights the following:

- capture/maintenance is performed once (no duplication) and the data is used in many thematic map publications (cartographic views);
- core systems are integrated and provide access to data for map product development;
- cartographic presentation is dynamically updated as maintenance updates occur in the primary geographic database - cartographic presentations are live views to the database;
- validation occurs at both the database management and cartographic production stages. Error notifications are triggered between the data management and product development environments;
- live spatial views enable data to be made available to customers in real time; and
- data dissemination is supported in a wide range of formats.

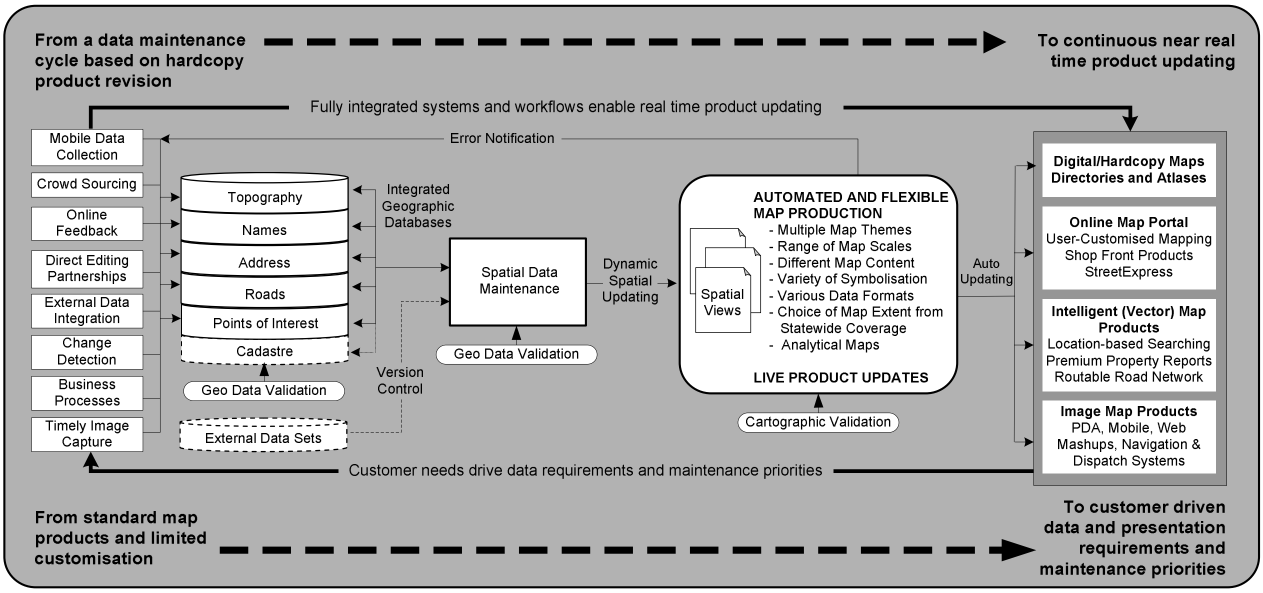

Figure 4: The Strategic Delivery System represents a conceptual view of the proposed business workflows (Landgate, 2009).

A solution consistent with the agency's enterprise architecture was essential to meet the short and long term needs of data managers and map producers. All data capture and thematic map production is conducted using the ArcGIS suite of software. In terms of longevity and sustainability, an off-the-shelf system was deemed preferable than building a new system from scratch. An evaluation of the software revealed that; (a) the data and product management processes could be managed using the available integrated workflows; and (b) Landgate's dynamic spatial updating methods were supported. The mapping tool set also enables geographic data to be enhanced for map publishing purposes without impacting on the underlying source data. Moreover, updates performed in the database are available to all product derivatives.

Quality control measures are supported and occur at key stages in the data and product management workflow. These validation tools ensure the integrity of feature attributes during data entry, the preservation of relationships between features, and spatial integrity (topological adjacency and connectivity). Cartographic conflicts occurring at the map level can also be detected automatically, although some additional scripts are required to manage integrity relationship with corporate systems.

4. IMPROVING DATA COLLECTION

Currently, it is necessary for Landgate's information compilers to actively source data. This is time consuming. The Framework enables this to change. The approach assumes the philosophy that – *customer-oriented products drive customer value-added inputs* - and this, in turn, will enable Landgate to build a comprehensive geographic knowledge-base for the future.

The most commonly used data collection methods are to; (a) interpret and extract geographic features from imagery using photogrammetric techniques; and (b) collect information from trusted agencies using manual business processes. The currency of geographic data is therefore, linked to the date of photography (often flown annually) or dependent on third party goodwill and resource availability.

More timely and sustainable data capture techniques are required to manage the lifecycle of geographic features. To achieve this, supply-chain business models are being developed to examine the way in which geographic features are collected, managed, processed and translated into map product representations. As the majority of these approaches are manual, the objective is to convert them to electronic data sourcing methods. Approaches include crowdsourcing, direct agency editing partnerships, external data integration, automated change detection and feature extraction methods, and mobile data collection.

- <u>Crowdsourcing:</u> Crowdsourcing is increasingly becoming an important part of today's spatial information business. Notification and validation of real world change can be achieved by tapping in to the collective intelligence of the general public and agencies. Landgate is developing an online map-based crowdsourcing tool and will run a pilot program with a small group of trusted partners. This solicited feedback mechanism is also an opportunity to gain a deeper insight into what customers want. It is anticipated that the service can be applied to a number of innovative community-based projects including asset management and social networking.
- <u>Direct Editing Partnerships:</u> Landgate is currently engaging with State Government agencies to assist in the maintenance of geographic data via direct database editing methods. The State's geographic data is used as a base for a number of agency applications. Rather than wait for quarterly incremental updates, trusted agency partners can make direct updates to the geographic database. Updates are then trickle

fed into agency replicated systems on a nightly, daily or weekly basis. This methodology has the potential for cross government savings and forms part of the State's Spatial Vision (under development), which seeks to reduce data duplication across the public sector.

- External Data Integration: Landgate has developed software that allows outsourced data capture/maintenance to be integrated automatically back in to the primary database. In the past, validation was conducted manually and this detracted from the savings accrued through outsourcing. Data can now be checked-in automatically.
- Automated Change Detection and Feature Extraction: Landgate continues to work collaboratively with Curtin University on computational change detection solutions; however no fool proof methods have been developed so far. The CRC SI2 is now taking a lead role in this area.
- Mobile Data Collection: Volunteer data collection is starting to gain momentum; the notion being to provide mobile editing opportunities so that staff and volunteer data collectors can collect data using a common operating picture/environment. This has potential for improving data quality through field verification. Landgate is trialling this technology for field valuations and subject to the pilots success, it will be rolled out for geographic data collection.

With this range of contemporary data collection techniques, the process of updating spatial data will be far more productive and timely than using existing techniques alone. In addition, sourcing geographic change using multiple avenues provides a mechanism for validating changes leading to improved data integrity.

5. DATA MANAGEMENT CONCEPTS

Under the SDPM Framework, Landgate's data and product management is based on the principle of *phenomena-cartography independence.* Geographic features are stored according to how they are observed in the real environment, independent of any specific application.

Cartographic (or product) views are simply cartographic representations derived from the source geographic data by way of a rendering process using the knowledge stored in the underlying data. The characteristics of features stored in the geographic database provide the core knowledge for deriving alternate cartographic representations. For example, the functional characteristics of a road, such as its surface material and class, are used to define how the road is symbolised graphically and hierarchically on a map.

This fundamental philosophy supports Landgate's vision of a multipurpose database from which various products can be derived. The way in which geographic data are modelled is key to enabling the flexibility to create products with a variety of themes and scales. The methodology contrasts with current techniques that are based on the storage of geographic features as sets of points, lines and polygons, according to their representation on a map.

With this traditional approach, geographic data is captured at various levels of representations to suit product-specific map scales and themes. These geographic representations are independently created and stored in a post generalised state. The approach creates data redundancy, as each geographic feature is stored more than once – one database object for each product representation required (Egenhofer *et al.*, 1994); (Timpf and Frank, 1995); (Kilpelainen, 2000). This creates duplicate data maintenance as each representation must be updated individually. It also makes it difficult to maintain representation levels without introducing inconsistencies in data currency (Egenhofer *et al.*, 1994); (Cecconi, 2003).

The approach adopted by Landgate supports a 'capture once - use many times' philosophy and is in line with the sustainability principles embedded in the SDPM Framework. As features are updated in the database, so too are their rendered product representations (see Section 6). This reduces the reliance on resources for data and product maintenance.

In applying this data and product management approach, not only must the completeness, currency and accuracy of data be of a high level, but the actual level of knowledge (attribute data) must be robust enough to cater for the various cartographic representations (symbolisation and map content) required. Whilst, the additional attribute knowledge required for theme and scale variability creates a data maintenance overhead; a knowledge-enriched data source is a valuable resource and has greater potential for diverse analytical and graphical applications.

A further objective of the SDPM Framework is to build a self-sustaining geographic data management environment. This will be achieved using two strategies. Firstly, the development of an extensible geographic data model that will provide Landgate with the agility to respond to customer needs for new data themes and feature characteristics (attributes); and secondly, automated schema evolution management techniques to manage changes when they occur.

Data schemas will always be subject to change – stemming from the need to rectify errors in the original schema, changing user requirements and advancements in knowledge domain requirements (Wachowicz, 1999). Landgate, in collaboration with Curtin University of Technology, is conducting research into methods for managing schema changes effectively. This is an important study - changes to the data model can have an adverse impact on spatial views if poorly managed. The objectives are to:

- ensure that the underlying data model is robust enough to withstand the addition, deletion and modification of data themes, attributes and relationships; and
- build algorithms to manage the change transparently without impacting the map products and applications in use.

6. AUTOMATED MAP PRODUCTION AND REVISION

A database-sharing and spatial (map) views environment is adopted for map production. This is a significant paradigm shift away from managing disconnected product data sets and enables sustainable business processes to be implemented through task automation.

Map production, in this paper, refers to the conversion of raw geographic data, stored in the primary database, to achieve cartographic presentation quality. The conversion process takes full advantage of GIS capabilities for cartographic generalisation, feature symbolisation, text placement, map page generation, formatting and marginalia, spatial indexing and gazetteer production.

Spatial queries are used to control thematic content. These queries act as a filter for each geographic theme and use the thematic characteristics (attributes) of features as query parameters. For example, roads may only be shown if they are classified primary and secondary and only if they have a sealed surface (Figure 5).

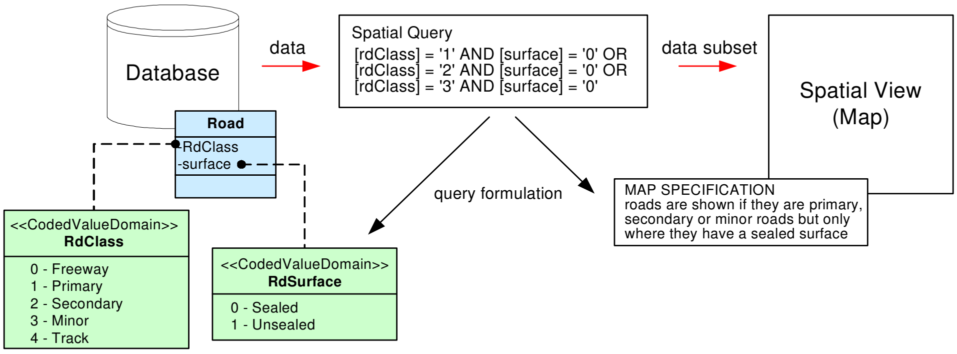

Figure 5: Process for specifying spatial queries to derive a cartographic data subset according to map specification (Arnold, 2006).

Cartographic enhancement (fine adjustments) is achieved using cartographic overrides. Whilst this changes the presentation of the features for map display, it does so without altering the underlying primary data. Similarly, rendering processes are used to alter feature geometry. For example, parks (polygons) can be collapsed to point feature representations, rivers can be smoothed and roads simplified[1]. These geometric changes are graphical manipulations only and are achieved without impacting the database geometry. This means that the database may be shared between multiple users, each having their own customised spatial view.

In a database-sharing and spatial views environment, spatial views are not separate digital files (in the traditional sense) but rather live windows to the underlying database. This means that updates performed in the database are also available (transparently) to spatial views.

The way an update appears in the spatial view is dependent on the dynamic spatial updating (DSU) rules specified in the view. These rules (or pseudo rules) govern the representation of features in the map product (view) and include the techniques, spatial queries, rendering processes and algorithms that are applied when a map view is created. Once applied these DSU rules persist in the product. When an update occurs in the underlying geographic data, the DSU rules associated with the product are triggered and the rules convert the update to the product-specific representation (Figure 6).

Cartographic validation is also integral to the longevity of the system and products. It ensures that data currency, at any point in time, is consistent across all products. Automatic cartographic conflict detection mechanisms are used to examine maps for illegible data clashes. These are corrected using the cartographic override functionality at the product level. All data errors (not cartographic clashes) detected in the product are fixed in the source data. In this way, the correct data is available immediately to all products derived from the source data, and this ensures consistent data management.

[1] Landgate uses a combination of data modelling, cartographic overrides and rendering techniques to implement cartographic generalisation.

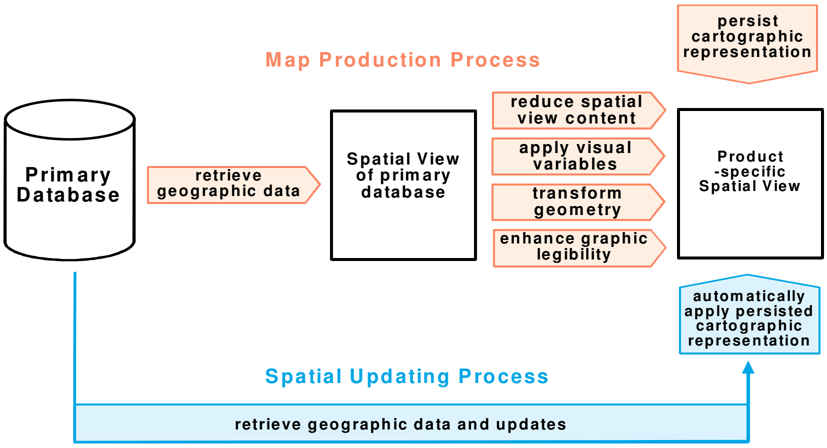

Figure 6: Cartographic processes applied during the mapping process, act as a pseudo rule-base for triggering updates (Arnold, 2006).

8. CONCLUSION

This paper presents Sustainable Data and Product Management business processes that will enable Landgate to continue to develop map-based products well into the future. With current map production processes, products are time consuming to produce, and this has caused delays in getting new products and services to market.
The SDPM Framework will implement streamlined workflows that enable real time product updating, reliable data quality controls and high quality outputs. This will free up resources and, in doing so, create a map production environment where the focus is on the development of new products and niche markets, rather than the maintenance of existing products. This will in turn create a more rewarding work environment.
The new workflows employ 'state-of–the-art' GIS and build on best business practices. A significant benefit for map production is the reduction in product maintenance costs achieved by eliminating data redundancy and duplicate data management. The potential for human error is also mitigated through task automation.
A key component of the Framework is to foster customer feedback and collaboration. Being able to work more closely with map users will allow Landgate to develop community focused products and services. Collaboration is also a mechanism to achieve timely data maintenance through crowdsourcing and direct editing opportunities. As a consequence, geographic data and products will be more up-to-date and currency will be consistent across all products and data sets. This will sustain public confidence in the delivery of government services.

REFERENCES

Andersen, R.S. (1997) *From CAD to GIS in Mapping Production* [Online], Available: www.esri.com/library/userconf/europroc97/5cartography/C4/C4.htm (29 July 2001).

Arnold, L. M. (2006) *Dynamic Spatial Updating: A Centralised Scale-independent Approach to Spatial Data Management and Thematic Map Revision,* PhD Thesis, Curtin University of Technology, Perth, Western Australia, pp

Beard, R. (1993) The Pros and Cons of Mapping with ARC INFO, *The Bulletin of the Society of Cartographers,* Vol. 27, No. 2, pp. 11-14.

Cecconi, A. (2003) Integration of Cartographic Generalization and Multi-Scale Databases for Enhanced Web Mapping, *PhD Thesis*, University of Zurich, Zurich, 155 pp.

Dobson, M.W. (1988) Digital Cartography in the World of Commercial Publishing, *Proceedings of 3rd International Symposium on Spatial Data Handling*, Sydney, Australia, August, pp. 17-19.

Egenhofer, M.J., Clementini, E. and Di Felice, P. (1994) "Evaluating Inconsistencies Among Multiple Representations", in: *Advances in GIS Research (Proceedings of the 6th International Symposium on Spatial Data Handling)*, Vol. 2, Waugh, T. C. and Healey, R. G. (eds.), Taylor & Francis, London, pp. 901-919.

Ertle, M. and Lauinger, H. (1995) Data-Based Map Production in Microcomputer Environments: A Professional Solution with the Nuages Software Package, *Proceedings of 17th International Cartographic Conference and 10th General Assembly*, Barcelona, Spain, September, pp. 1452-1459.

Kilpelainen, T. (2000) Maintenance of Multiple Representation Databases for Topographic Data, *The Cartographic Journal,* Vol. 37, No. 2, pp. 101-107.

Landgate (2009) Sustainable Data and Product Management Strategy, Unpublished Report, pp 29.

Leskinen, T. (1995) Publishing Topographic Maps using Postscript Page Description Language, *Proceedings of 17th International Cartographic Conference and 10th General Assembly*, Barcelona, Spain, September, pp. 1145-1149.

Peck, S. (1995) *Colour Publishing of ArcInfo Maps - Trial, Error and Success* [Online], Available: http://www.esri.com/library/userconf/proc95/to150/p146.html (31 July 2001).

Sena, M. (1991) 'Map Publishing', In *Computer Graphics World*, Vol. 14, pp. 75-77.

Timpf, S. and Frank, A.U. (1995) A Multi-Scale Dag for Cartographic Objects, *Proceedings of ACSM/ASPRS Annual Convention and Exposition, Autocarto 12*, Charlotte, North Carolina, USA, February-March 1995, pp. 157-163.

Wachowicz, M. (1999) *Object-oriented Design for Temporal GIS*, Taylor and Francis, London, UK. 118 pp.

BIOGRAPHICAL NOTES

Lesley Arnold completed a PhD at Curtin University of Technology and was awarded the D.B Johnson Award for Excellence by the Spatial Sciences Institute. Her research examined an integrated solution to data management for map publishing and automated map revision. Lesley currently works at Landgate as Manager Location Products and Services coordinating a number of geospatial information projects. She also consults internationally for Landgate in the fields of Mapping, Land Information Systems, Spatial Data Infrastructures and Education. Lesley is currently Adjunct at Curtin University of Technology and is co-supervising PhD candidates in the fields of Schema Evolution and High Water Mark Determination.

CONTACTS

Dr Lesley Arnold
Landgate
1 Midland Square
Midland, Western Australia
Tel. + 61 8 9273 7144
Fax +
Email: lesley.arnold@landgate.wa.gov.au
Web site: www.Landgate.wa.gov.au