Supporting Dynamic, Evolving and Emerging Land Information

Geoff HAY, New Zealand

Key words: Land Administration, Mobile Technology, Cloud Computing, Computerization, Temporal Data Model

SUMMARY

Land administration situations that feature highly dynamic information and rapidly evolving and emerging understandings and practice, require appropriate computerized representation and implementation that supports variety, velocity, and evolution of data, its meaning, and the practices associated with its treatment. Such situations include land administration jurisdictions that are emerging or recovering from disaster or conflict, moving from manual to digital practice, or that feature highly dynamic and evolving extra-legal or secondary arrangements that are difficult to harmonize and resolve with state sanctioned or conventional treatments. The use of mobile technologies and publicly contributed information and claims regarding land and tenure provide a useful and timely means by which such data may be sourced in these situations although such means introduce additional potential for corruption and for ambiguous, conflicting, and imprecise data.

This paper describes an approach to representation that addresses the above concerns. The temporal nature of the shared and agreed understandings relating to the organisation of land, land tenure, and administration, and the specification, implementation, deployment and maintenance of computerized land information (LIS) systems is examined and an amendment event-based bi-temporal model is developed. The model well supports the faithful recording of history and tenure reality and a cloud implementation that combines aspects of social networking and contributed knowledge, and support for large variable datasets and analytics. The model has particular relevance to the treatment of customary, extra-legal and secondary associations to land, to dynamic, emergent and evolving concepts of LA, and to such concerns as recovery from disaster and sustainability over time. In addition, the model and treatment provide representational support for adjudication processes and automated determination of claims.

Supporting Dynamic, Evolving and Emerging Land Information

Geoff HAY, New Zealand

1. INTRODUCTION

In land administration, the faithful recording of history underlies the formulation of manual paper-based record keeping practice. Such practice relates to the fundamental requirement that the full history of change is recorded and that information is not overwritten or otherwise lost through the recording of change. The time stamping of documents and other artifacts (such as geometric representations of parcels) relating to land transfers or mutations is the conventional means of recording change in cadastral information stored in databases.

Database queries may be used to retrieve the set of documents relating to a particular parcel, or owner, and for a particular time-period. If these documents are scanned images then the information content of these documents may not be available for querying unless the content has been extracted into attributes and tables and appropriately indexed. This representation of time is suitable for situations where there are only a few kinds of documents (or instruments) that may be applied to a cadaster and where the kinds of documents do not change often.

A more compressive system explicitly represents the content of documents although this introduces a range of issues not the least of which is the difficulties associated with evolving databases which explicitly record document content as attributes and relations. Generally these attributes relate to interest holders and land owners, the parcels and the various kinds of relationships that may exist between them, and to land transactions that change attribute values.

2. DYNAMIC, EVOLVING AND EMERGING LAND INFORMATION

In highly dynamic cadasters, change in data values (e.g. ownership or boundaries) may occur at such pace (velocity) that it is difficult for paper-based systems to remain up-to-date and therefore authentic. Digital solutions are essentially required in these situations especially where corruption is a problem. A cadaster that is not up-to-date and has lengthy transaction processing times invites corruption and potentially fills the courts with land-related disputes. Conventional digital land information systems may struggle to remain up-to-date in countries that feature large numbers of small plots of land and large numbers of land transactions.

Cadasters that are rapidly evolving in response to new policy, the requirements of economic development, population growth and/or movement, or in recovering from disaster, may have difficulties with conventional digital systems that are difficult to evolve in response to changing conditions. Typically, the more comprehensive a digital cadastral database and software application is the more complex and costly it is to evolve. Even the simplest of changes (e.g. adding new attributes) requires change to schema, to existing data, and a flow-on change to software and other artifacts (e.g. document templates, queries).

Emerging information, such as that resulting from registration activities or crowd-sourcing, may include not only the value data associated with boundaries or tenure, but also the practices, language, and the particular tenure types and organisation of land and its administration. The databases and software that support emerging information may not be aware in advance of the kinds of information that will emerge. Generally, the more specific a schema[1] is the less it provides for variation in the kinds of information it can support and the way in which requirements might evolve over time. Although less specific (or more abstract) logical schema can support more variation in the kinds of information, this is often at the expense of querying, an example of less specific schema is the social tenure domain model (STDM) (Augustinus, et al, 2007). Essentially queries become less reliable (or more complex) because there is less certainty regarding the kinds of things stored in the database.

Dynamic, evolving and emergent concepts of information are inherently temporal concerns. Logically, in order to properly deal with these concerns, a temporal logic and organisation of information and supporting infrastructure is required. Ultimately, the definitions of a domain, which might derive from laws, civil codes, oral traditions, or even crowd-sourced information, also follows a dynamic, evolving and emergent temporal conception. A temporal model for information and implementation architecture that addresses these temporal concerns and which relates to a ubiquitous spatio-temporal data infrastructure was defined by Hay (2014).

The basis of the proposed architecture is the general concept of an *instrument* to serve as a unit of domain specification and system implementation. The concept of *prescriptive instruments* was introduced to represent generally, the shared and agreed understandings that specify the form and practice of a domain. The evolving set of prescriptions defining a domain drives the evolution of the requirements for information recording and information systems implementation. The prescriptive variation that occurs across domain contexts is the conceptual analogue of the variation that occurs across time (evolution) since the impacts on requirements are essentially the same. The thesis focused on a temporal solution to the problems of variation that impact the longevity and relevance of information systems.

Individually, prescriptive instruments encapsulate a portion of reality that is a temporal statement regarding the state of a domain understanding. A prescriptive instrument appends to a domain understanding, terminology and meaning, and the structures and rules associated with the content and form of both its data, and its practices. The legal activities and practices of a domain which relate to the inscription and retrieval of information and the processing of formal documents are

[1] In this paper, the term ‘schema’ is used to refer to artefacts that define in advance the specific local models of information for use in creating databases and software.

represented with the concept of *transaction instruments* and a process-based perspective. Transaction instruments represent the basic unit of information system specification and deployment and the organizational basis for the other prescriptive elements. Transaction instruments are implemented as software processes or workflow models and relate well to the idea that a land administration is essentially a set of inscription and retrieval processes.

The instrument-centred software architecture defines an organisation of system components and treatment of information which replicates an evolutionary model of system specification and development through its support for the incremental implementation of prescriptive and transactional instruments. The architecture specifies separate perspectives of information organised around a temporal model of information to underpin an evolving and emergent paradigm for information systems. The logical storage of information is based on an append-only temporal model of data rather than a structural representation of the domain allowing the structures and meaning of information to vary independently of implementation.

Domain understanding is captured in a layered ontology (*semantic perspective*) that logically centralizes the terminology and the semantics associated with stored time-referenced data. The transaction instruments are implemented as software processes that orchestrate and organize human interaction with information system components and applications to achieve inscription and retrieval processes. The *process perspective* coordinates and controls the interaction between the information perspectives, the application of rules and constraints, the determination of data structures, and forms the basic integration mechanism between general and specific software components and services. A process-based approach also supports the separation of the spatial dimension of data and allows reuse of existing GIS technology to support the geometric or purely spatial aspects of data. The incremental definition and deployment of process implementations forms the basic unit of system evolution and emergence.

The approach supports a solid foundation for maintaining and retaining data to provide authentic and reliable evidence of record keeping activities. An append-only storage model for data and its logical conclusion of amendment-based data update based on the appending of facts to a data store captures the idea that the complete truth about change can be recorded. An event-based logical model serves as the organizational basis for the logical storage of facts and amendments and has the advantage of being highly reusable in contrast to its relational counterpart which models specific local domain knowledge. In addition, an event-based bi-temporal model allows for the complete recording of history including post-active and retrospective updates, and for the state of a domain to be reconstructed for any point in time, and for any temporal context.

3. TIME

The definitions or prescriptions that define a domain relate to a temporal formulation that has two important features: prescriptions are added and may reference existing prescriptions; and, existing prescriptions can never be deleted or otherwise removed from history although they may be amended, replaced, or superseded by new prescriptions. This append-only amendment-based strategy forms the logical basis for a temporal model for information that can be applied to both data and metadata and which fully supports the faithful recording of history.

One reason why computerized support for the faithful recording of history has not been addressed in a comprehensive manner is that it requires that all data are retained forever and that no information is lost through continuous updating. The traditional concerns associated with storage and processing capacity are less-relevant in these days of virtualized and massively parallel computing. However, in retaining historical information it becomes necessary to also retain the history of and support for evolution in the logical schema of recoded information in order to maintain support for existing stored information. The essential question is: what should be done as the schema of information changes over time? Conventionally, schema evolution is achieved through versioning of schema and/or processes of data migration where existing data are migrated to new schema, and/or schema negotiation where several schema versions might be considered in search and retrieval operations. These concerns have been recently reignited with the debate over crowd-sourcing and participatory approaches to land information collection where questions relating to authenticity and the applicability of imposed schema have been raised (Laarakker et al, 2016).

Interestingly, the issue of schema evolution is, in practical terms, the same as the issue of schema variation which is where similar phenomena are recorded under differing conceptions. For example, schema differences due to design decisions by different people that result in different names for the same concepts and other kinds of schema heterogeneity. The variation in schema that occurs across space due to variations in laws and other prescriptions also results in such schema differences. The conventional approach to dealing with schema variation is through standardization, for example the land administration domain model (LADM) (ISO 19152) standardizes a domain schema for use in integration of land administration data. Although standardization addresses schema variation problems, it essentially assumes the underlying agreements do not change over time and can be fixed forever. This effectively limits the schema to only those concepts which are able to be fixed. Standardization does not address evolution despite the fact that evolution and variation are 'two sides of the same coin' (Roddick et al. 2001).

A practical implementation architecture for schema that can evolve without requiring intervention (data migration or schema negotiation) as usage evolves and new concepts are included but which can, at the same time, support varied requirements and data structures has been proposed (Hay, 2014). Such support requires two essential ingredients: its schemas are made up of semantic descriptions of concepts and roles defining elements that *may* be used or referred to in data rather than defining in advance what *must* occur; and, the data themselves should conform to a logical model that does not subscribe to any particular domain conception. That is, the logical model for data and the recorded data are not dependent on any specific domain understanding. This can be achieved by using a logical model for data based on concepts of time rather than domain understanding, and the data themselves refer to their specific domain understanding via references to semantic descriptions held in a semantic repository. Such a logical model does not need to be modified in order to be transported across context (variation) or time (evolution).

This definition of a schema differs from conventional schema in that those structural elements that are likely to vary in form across context or time (such as the structure of particular documents or the specific structures made up of relationships and entities in a domain) should not be included. This is so that schema can remain relevant over time; and, schema must be *append-only* in that new roles and concept descriptions can be added over time but cannot be removed. This concept of schema definition and evolution mirrors that of the underlying prescriptions from which the descriptions which make up the schema are derived. This concept of schema thus relates more to a catalog of semantic descriptions rather than a domain model and database definition.

4. LOGICAL MODEL

The logical model to underpin storage of data defines a set of abstract entities based on a bi-temporal logic that explicitly records change in geographic phenomena and retains fully the history of change. The model implements a multi-threaded linear time recording the attribute history of entities and their interconnections. The abstract temporal concepts of ‘Identity’, ‘Process’, ‘Service’, ‘Event’, ‘Fact’ and ‘Amendment’ together with the semantic concepts of ‘Concept’ (or ‘Class’), ‘Property’ (or ‘Role’), the use of internet protocols, and the resource description framework (RDF) make up a means by which any domain phenomena, including their semantic, spatial, and temporal dimensions, may be described.

The abstract concept of an *Identity* is used to represent any domain phenomena which can and should be uniquely identified and which is worthy of recording. The Identity concept serves as a container for attribute information relating to identity instances including information defining an identity’s actual real-world types. Changes to the data are achieved with *Events* which define the transaction-time associated with the *Facts* that they apply to identities. An identity is created by an initial event and subsequent events may add new attributes and values or modify existing attributes and their values. An identity (and in fact an entire database) is thus composed of a set of appended events which are well-formed snippets of information comprising sets of facts. Once an event associated with an identity is appended to a database it cannot be removed and it cannot be modified. This implements the append-only concept. In order to implement change an amendment-based strategy similar to that used in accounting practice and extending the *amendment vector* approach described by Langran and Chrisman (1988) for vector data was employed.

In that approach, amendment vectors defined links from child to parent polygons in a spatial database allowing the history of subdivision and merge operations to be recorded and for historical maps to be reconstructed. This idea was extended with bi-temporal times-tamping where both transaction-time and valid-time are recorded allowing retrospective and post active updates, and, to include all attributes so that the history of aspatial as well as spatial change can be recorded and state reconstructed. Reconstruction of an historical state is simply a matter of eliminating those events or facts which relate to timestamps that occur after the time for which the reconstruction relates. Note that every query thus requires a temporal context defining whether retrospective or retroactive updates are to be considered when reconstructing state.

The concept of change was implemented with the abstract concept of *amendment* and amendment logic defining concepts such as *supersedes, retires, amends, expires* relating to the way in which individual attributes may be changed. A fact imposed by an event may change the value of an existing attribute by referencing an existing fact (using an *amendment*) and supplying a new value. Facts may add new attributes to an information identity and remove attributes from use. The attributes themselves are made up of references to semantic descriptions of roles and may reference semantic descriptions of types, references other artifacts (such as documents) or literal value data.

The idea that even incorrect information that exists in the database is never lost in the face of corrective amendment is an interesting side-effect of the bi-temporal model. This fits well with the faithful recording of history where, for example, an incorrect value placed into the data may be later corrected with a corrective amendment that does not hide the fact that an incorrect value existed and was considered correct for some period of time. Such information may be exposed when queries are expressed in terms of transaction or valid time.

Ontology is used to record separately explicit descriptions of properties and types. Essentially, the abstract logical model employs RDF and OWL in order to separate the concerns of time from the specific meanings and understandings of the recorded instance data thereby completely separating the handling of domain data from their specific domain meaning. The instance models which record the values and history of real world objects are sets of events made up of three-part RDF statements composed of navigable URI references to specific information such as references to geometry held in spatial databases, semantic descriptions held in ontology, scanned documents held in a document database, or software modules. Statements may also include actual literal values where this makes sense, for example, time stamps.

The above combination of ontology and a logical data model implemented with the same technology (i.e. RDF/OWL) provides for a ubiquitous database model in which *"anything can be said about anything"* however, while this provides for an easily reusable database, it provides little in the way of control relating to updates and transaction. This is achieved with the *processing perspective* which implements the services and processes that control data integrity and consistency, and implement generic and user specific inscription and retrieval processes. Importantly, the architecture requires that it is only through the action of processes that information may be updated (through authentic append processes). The process perspective is also where queries are implemented. In this approach, queries are implemented as ordered organisations of generic tasks that incrementally achieve the query goals. Query tasks relate to spatial, semantic, temporal perspectives of information in a similar fashion to the three domain model proposed by Yuan (1996, 1999) which also separates domains of space and time to address the problems of complexity associated with changing spatial and topological properties of features, as well as time-varying aspatial properties. The advantages of this approach relate to reduced complexity for users due to the separation of query concerns into discrete tasks that may be reused across query processes. A general implementation of the process perspective was described in Hay & Hall (2011) and a range of spatio-temporal query examples are given in Hay (2014).

The temporal model and architecture allow for the definition and unique identification of software modules, and for these to be created, deployed, versioned, composed into applications, and for these to be referenced within data. This provides for an evolving and evolvable system concept. The definition of these software artifacts derive from the same prescriptive definitions as the semantic and processing definitions. The software modules implement specifics of the domain that are variable or that evolve over time, for example, the structures of data may vary across records, and the form and attributes of documents may change over time. The theory provides for the idea that a software service that is responsible for creating a particular kind of document (its version, format, and content), can also read that document and can provide services corresponding to that particular kind of document.

5. ADVANTAGES

The architectural model relates to support for the translation of domain prescription into computerized implementation, specifically, the translation of prescription into software and databases supporting land administration. The process of implementation follows a pattern of translation of prescription into perspective artifacts. Terminology including classes, concepts and roles are translated into semantic descriptions if they do not already exist. Transaction instruments are translated into processes either as software applications or workflow models that can be executed as needed. These processes refer to and reuse generic inscription and retrieval services and services that implement generic or custom functionality such as advanced analytics and fraud detection. Specific fixed phenomena such as formal document types are translated into services that implement their definition and provide operations specific to their handling and content. For a particular jurisdiction, this means that development of software code and process models is only required for that which is specific to that jurisdiction although these may refer to existing services to provide the generic aspects. This represents advantages in that there is less need to 're-implement the wheel' (in particular the temporal concerns) and there is less need to be concerned about infrastructure aspects – especially if cloud-based infrastructure can be reused as a service.

The temporal model and architecture are well aligned with a cloud-based deployment and big data treatments for information. This is mainly due to the fact that data are held under a ubiquitous logical model that can be efficiently utilized by cloud and big data processing capabilities. Cloud-based deployment potentially allows a more cost effective provisioning of infrastructure especially in support of the potentially large amounts of data and processing requirements associated with the faithful recording of history including the notion of an easily evolvable system that does not require re-engineering of implementation or data in the face of constant requirements change. In addition, the potential for the provision of advanced services that might otherwise be unaffordable, especially as requirements for land information evolve and become more sophisticated cannot be ignored.

The power of big data approaches in retrieving and analyzing information cannot be ignored and these can only improve with the inclusion of semantic and rule-based inference techniques. These capabilities may prove to be extremely valuable in processing and adjudicating crowd-sourced and contributed information.

6. CONCLUSION

This paper described the potential advantages of a proposed temporal model to underpin the storage, retrieval and processing of land related data, and the definition, implementation and deployment of software to support the administration of land. The temporal logic and the model concepts were derived from the temporal properties of the prescriptions (e.g. laws, codes, oral traditions) that define specific land administration domains (Hay, 2014).

The approach includes a logical data model suitable for cloud-based deployment and big data analytics, a logically centralized repository of semantic description of classes and roles, an architectural and philosophical requirement for a basis in processes as a formal means to control updates and retrieval, and a service oriented architecture as a means towards component-based versioning and system evolution and to provide for evolution in data structures and document forms relevant to the faithful recording of history. In addition, the approach provides for the full retention of the history of schema change and usage. Essentially, the approach allows for schema to vary and evolve and for exact schema to be determined from usage with the data after the fact rather than requiring all to be known in advance. This greatly supports concepts of emergence, variability and evolution.

In support of emerging information that arrives from less trusted sources the power of big data analytics should not be underestimated especially when this is in combination with semantic descriptions that underpin terms referenced within the data. This potentially improves support for automation in determination, decision making and adjudication processes through improved capability to consider and range over similar but different phenomena, and infer logical consequence from incomplete or conflicting information. The potential for such processing to support anti-corruption and fraud detection measures could be further investigated.

Finally, the approach defines all the elements of a ubiquitous data infrastructure extending the concept of a spatial data infrastructure (SDI) into a spatio-temporal data infrastructure. Such an infrastructure supports the concept of land administration ‘as a service’ providing for cloud-based, and potentially multi-tenanted shared land administration and management services while still maintaining local character and requirements. Such arrangements potentially achieve a range of cost savings and benefits especially in less well-resourced jurisdictions. These include costs savings associated with provision of computing infrastructure and the development, deployment and maintenance of software and databases. Potential benefits include improved authenticity of land information, faster deployment and maintenance of systems, faster land transaction and land registration processing times, and the cost-effective provision of advanced capabilities. In terms of responsible land administration (Laarakker et al, 2016), crowd-sourcing and adjudication services and processing, the semantic capabilities provide a potential to improve the authenticity and usefulness of such information in protecting interests, rights and tenure for land occupiers and users.

REFERENCES

Augustinus, C.; van Oosterom, P.; van der Molen, P. & Lemmen, C. The Social Tenure Domain Model --- Design of a First Draft Model FIG Working Week, 2007

Hay, G. C. (2014). Architecture for instrument-centred land administration applications (Thesis, Doctor of Philosophy). University of Otago. Retrieved from http://hdl.handle.net/10523/4985

Hay, G. C. & Hall, G. B. (2011) Implementing Open Source Software for Land Administration Processes in Developing Nations FIG Working Week 2011 - Bridging the Gap between Cultures

ISO 19152 Draft International Standard ISO/DIS 19152 Geographic information — Land Administration Domain Model (LADM) ISO/DIS 19152, 2011

Laarakker, P., Zevenbergeb, J., Georgiadou, Y. (2016) Land Administration Crowds, Clouds, and the State. In Zevenbergen, J., De Vries, W., Bennett, R., (Eds). Advances in Responsible Land Administration (pp. 91-114), Taylor & Francis Group, Boca Raton.

Langran, G. & Chrisman, N. R. A framework for temporal geographic information Cartographica: The International Journal for Geographic Information and Geovisualization, 1988, 25, 1-14

O'Sullivan, D. (2005). Geographical Information Science: time changes everything. Progress in Human Geography 29, 6 pp. 749-756

Roddick, J. F., Grandi, F., Mandreoli, F., and Scalas, M. R. (2001). Beyond Schema Versioning: A Flexible Model for Spatio-Temporal Schema Selection. GeoInformatica, 5(1), 33–50.

Yuan, M. Use of a Three-Domain Repesentation to Enhance GIS Support for Complex Spatiotemporal Queries Transactions in GIS, Blackwell Publishers Ltd, 1999, 3, 137-159

Yuan, M. Temporal GIS and spatio-temporal modeling, Third International Conference on Integrating GIS and Environmental Modeling, NCGIA, 1996

BIOGRAPHICAL NOTES

Geoff is an expert in land administration and land information technology. His focus over the last decade has been on land administration practice in both developed and developing nations and the development and deployment of technology solutions to the many and varied problems related to sustainable and secured land administration and land tenure. Geoff has completed a PhD in land administration and information science at the School of Surveying, University of Otago in New Zealand and his thesis relates to the science of spatio-temporal land information systems. Geoff currently works for the Land Administration group at Trimble Navigation. He is a firm believer in the economic benefits of authentic and sustainable land administration and the need for intelligent and evolvable information systems to support responsible land administration.

CONTACTS

Geoffrey C. Hay
Trimble Solutions
11-15 Birmingham Drive
Riccarton 8024
Christchurch
NEW ZEALAND
Tel. +64 3 39635524
Fax +64 3 39635417
Email: geoff_hay@trimble.com
Web site: www.trimble.com