DIGITAL PRODUCTIVITY FLAGSHIP
www.csiro.au

A Data Specification Framework for the Foundation Spatial Data Framework

Paul Box[1], Bruce Simons[1], Simon Cox[1], Stephen Maguire[2]

[1] CSIRO Land and Water Flagship

[2] Zicomi Systems

A report for the Australian Government Department of Communications

Citation

Box P, Simons B, Cox S, Maguire S, (2015) A Data Specification Framework for the Foundation Spatial Data Framework. CSIRO, Australia. CSIRO

Digital Productivity Flagship | Land and Water Flagship

Copyright and disclaimer

Important disclaimer

Contents

Figures

Tables

Acknowledgments

This report was commissioned by the Department of Communications as part of the Foundation Spatial Data Framework Initiative (FSDF). The authors would like to thank the Department for supporting this work and acknowledge the contributions of the following individuals from the Department in developing this report: Jamie Baker, Brian Sloan and Tim Neal.

In addition, the authors acknowledge the contribution of the following individuals and organisations to the development of information modelling methodology and tools through project funding.

Bureau of Meteorology for co-investment (through WIRADA) in the development of Solid Ground information modelling toolset and registry. This capability is being used as the basis for the development of the model registry described in this document. Geoscience Australia for supporting FSDF modelling activities, and for inputs to inform the development of the data specification framework. Simon Costello, Dmitar Butrovski, Graham Logan, Andrew Whiting and David James. Marcus Roy (NICTA) for contributing to the FSDF model development for the Administrative Boundary and Transport Theme. Marcus Blake (Australian Bureau of Statistics) for contributing towards the development of the FSDF Administrative Boundaries models.

Finally, the authors would like to acknowledge the INSPIRE initiative and its critical role in defining data specification methodologies that have provided an invaluable resource to inform FSDF data specification efforts.

Executive summary

The purpose of this report is to present the rationale for and design of a model driven data specification framework to support the Foundation Spatial Data Framework (FSDF). The report: articulates the rationale for spatial information modelling, placing it in the context of the geospatial information supply chains; describes the modelling process; and presents a proposed FSDF data specification framework and key recommendations for its implementation.

Foundation Spatial Data Framework

The FSDF[1] is an ANZLIC initiative that aims to deliver national coverage of the best available, most current, authoritative source of standardised and quality controlled foundation spatial data for Australia and New Zealand. To achieve desired outcomes, the FSDF initiative must address a range of interwoven technical and social challenges caused by the fragmented and heterogeneous production, management, supply and governance of geospatial data across multiple levels of government. The most critical challenge is the integration of a patchwork of data sources from all levels of government with different structures and semantics (or meaning), developed under different business contexts, into a coherent suite of maintainable interoperable national products. To address this challenge, a data specification framework underpinned by model driven approaches is proposed.

Standards based modelling

Modelling the information used by an application is the most important element of the data specification process. Formal modelling provides significant benefits when developing interrelated spatial data products, and so information models are the foundation of the FSDF. Information models will enable efficient production, maintenance and use of suites of interoperable, standards based foundation spatial data.

The 19100 series of international standards developed by ISO/Technical Committee 211 provides a conceptual modelling framework for geospatial information. These standards provide a framework within which information models can be developed for different application domains in a consistent manner. For example, the INSPIRE data specifications[2] are developed using ISO standards as a framework which results in full interoperability between data of the same theme from different providers, as well some level of interoperability between data across different themes. Data specifications, such as those developed by INSPIRE, are standards that can themselves be promulgated in a specific community. This is the case for the models that will be developed under the remit of the FSDF as they will act as standards for data product development and data exchange.

Developing FSDF data specification using the ISO 19100 framework enables:

- Quality and fitness for purpose – the opportunity to update existing or develop new products that meet end user needs in the context of emerging technology paradigms;
- Interoperability – achieve interoperability between products by specifying the things that need to be same across the whole of the FSDF as well as within each theme;
- Harmonisation - improve coherence between data sets within and across themes; and

[1] http://www.anzlic.gov.au/foundation_spatial_data_framework

[2] See Technical Guidelines Annex I http://inspire.ec.europa.eu/index.cfm/pageid/2

- Efficiency – improvement in supply chains to reduce production and use costs, and address common challenges which are ubiquitous across FSDF and underpin common solutions.

Geospatial information supply chains

FSDF is concerned with developing a coherent framework of foundation spatial data. In this context, the value of modelling is in enabling the evolution of existing supply chains to more efficient approaches, including increased automation, machine to machine interactions and the delivery of smarter, timelier geospatial products via Web services. The report identifies five important supply chains patterns for geospatial information supply:

1. anarchic (or point to point) – direct producer and user interaction;
2. centralised - centralised production of data by a single organisation;
3. aggregated - aggregation and integration of data by an intermediary;
4. brokered - centralised broker transforms heterogeneous supplied data to a common form; and
5. federated - federated data supply using common community models.

The patterns are distinguished primarily by the location and timing of the transformation of supplied data from source structure and format to that of the final delivered 'information product', using an application schema (or data model). The analysis of the supply chain patterns identifies the (largely hidden) and significant costs borne by end users of modifying products which are not fit for purpose. The analysis also highlights the critical role of information models to enable effective design and delivery of fit for purpose products through the range of supply chain models. Based on this analysis, the case is made for investment in a data specification framework.

The information modelling methodology

This report describes a modelling approach that is used for the FSDF. This modelling approach, which places the structural definition of information at the centre of the design process, is known as the Model Driven Architecture (MDA) approach, and has been used successfully in information technology for many years. The modelling approach is essentially a standardisation process that leads to the development of data products or data exchange schemas that can be adopted by a community to deliver consistent data. This report describes key modelling concepts implicit in the use of ISO 19100 series and Open Geospatial Consortium (OGC) standards, and underpins spatial information modelling efforts. The types of models used and an overview of the arrangements of those models in the FSDF data specification framework are also described.

The data specification framework

A data specification framework, parts of which have been implemented, is presented. The framework comprises three interrelated components:

- a suite of modular and interdependent models, and controlled vocabularies that define foundational spatial data and the way in which it is used;
- modelling tools and processes – the tools, systems and processes used to create, access and exploit models and vocabularies; and
- model and vocabulary governance - the roles, processes, rules and mechanism for the governance of models and vocabularies throughout their entire lifecycle from creation, publication, use and retirement.

The driver for development and operation of the data specification framework is the development of application schemas that will be used to refine existing and develop new national foundation products. The report notes that the implementation of the proposed data specification framework, will only realise benefits if the framework is used to support the development of national foundation data products. There is a cost associated with the operation of the data specification framework and further investment in developing FSDF information models, implementing model governance, and developing modelling tools and processes, is not recommended if models serve simply as documentation of existing data products.

1 Introduction

1.1 Purpose

The purpose of this report is to explain the rationale for a geospatial data specification framework and describe a model driven approach to data specification to support the Foundation Spatial Data Framework (FSDF).

The report is aimed at a number of audiences:

- **senior executives** – to explain the rationale and benefit of modelling, in the context of the development of a national information infrastructure - the FSDF (sections 1 and 2);
- **technical managers** – to articulate the value and process of modelling from both an enterprise (i.e. single organisation) and system of systems i.e. the FSDF perspective (all sections);
- **information modellers** - to place modelling activity in a broader context which will enable a clearer understanding of the rationale for and role of data specification (sections 2, 3, 4 and 5); and
- **database, data product and service implementers** – to understand the role of data specifications for supporting the design and implementation of related products and services (sections 3, 4 and 5).

This report is intended primarily for those engaged in the FSDF. However, much of the content is relevant for other parts of the spatial information community or more broadly for other non-spatial government data initiatives.

1.2 Scope

The report describes the rationale for spatial information modelling and the role of information modelling in the context of geospatial data supply chains. This supply chain analysis excludes data collection concerns. The report then describes key modelling concepts and the FSDF modelling process. A proposed data specification framework is then outlined. The report does not provide a detailed coverage of the data specification framework as this will be articulated during implementation. However, it does provide recommendations for next steps.

1.3 The Foundation Spatial Data Framework

1.3.1 OVERVIEW

The FSDF[3] is an ANZLIC initiative that aims to deliver national coverage of the best available, most current, authoritative source of standardised and quality controlled foundation spatial data for Australia and New Zealand[4]. Foundation spatial data provides the basic data infrastructure within which richer applications can be implemented. ANZLIC envisions foundation spatial data as a ubiquitous part of activities across all sectors of both the Australian and New Zealand economies.

The FSDF has been conceived as a national approach to enable access to and evolution of national foundation spatial data. It will provide a common reference for the assembly and maintenance of Australian and New

---

[3] http://www.anzlic.gov.au/foundation_spatial_data_framework

[4] The term 'geospatial' is used throughout this document to refer to both geographic data i.e. data that relates to locations on the Earth's surface and to 'spatial data' that describes the relative position of objects in space.

Zealand foundation level spatial data in order to serve the widest possible variety of users. The FSDF represents a data and user demand centric approach to developing national Spatial Data Infrastructure (SDI).

As noted by Drew Clark, the ANZLIC Chair, the key benefits to be realised through implementation of the FSDF are "improving supply chains, realising efficiencies and reducing the duplication of effort in the Australian, state and territory governments" (ANZLIC—the Spatial Information Council 2014). Ultimately having more accurate products that meet user needs and are created in a timely way will lead to better decision making and result in better outcomes for the Australian and New Zealand public.

The FSDF groups foundation spatial data into the following themes:

- Geocoded Addressing;
- Administrative Boundaries;
- Positioning;
- Place Names;
- Land Parcel and Property;
- Imagery;
- Transport;
- Water;
- Elevation and Depth; and
- Land Cover.

User consultations have identified priority datasets for each theme, and theme profiles have been completed that provide a description of each theme and identified data sets, together with use cases for the data. In parallel with the technical work, FSDF governance arrangements and a policy framework have been developed. Three-year road maps have been devised for each theme, which identify future goals and plans to resolve gaps, evolve datasets and delivery mechanisms. Work plans are articulated around four areas of focus as shown in Figure 1. A work plan for a 'cross cutting theme' has also been developed that addresses data specification and harmonisation priorities.

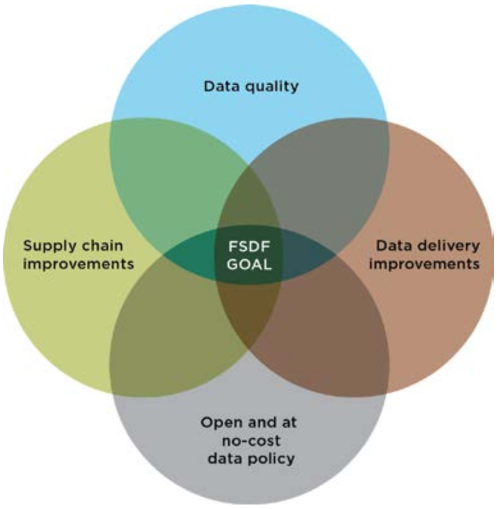

Figure 1 Evolution of FSDF datasets: Areas of focus (ANZLIC—the Spatial Information Council 2014).

1.3.2 GEOSPATIAL DATA PRODUCTION

Much geospatial data is developed and delivered in organisational or domain silos that are isolated from each other. For example, separate supply chains may exist for data related to the same phenomenon at different scales e.g. data about roads are collected via separate supply chains: national road data set at 1:1million scale; and state data sets at 1:50,000 scale. Furthermore, representations of the same phenomena are duplicated in different domains and are not connected with each other e.g. Local Government Areas (LGA) are represented by the Australian Bureau of Statistics (ABS) census geographies and duplicated in in state gazetteer records.

In Australia, data is geographically fragmented due to the federated nature of Australian government and the production of geospatial data by organisations at three levels of government.

This results in multiple representations and identifiers for the same real world objects in data, causing confusion and cost for end users needing to reconcile different views. An example is given in Figure 2 which depicts the identifiers for the suburb of Leichhardt in the New South Wales and Australian national gazetteers, and the Australian Statistical Geography Standard (ASGS) as an approximation of the suburb of Leichhardt. Furthermore, Leichhardt has an identifier in Geonames, a crowd sourced global gazetteer.

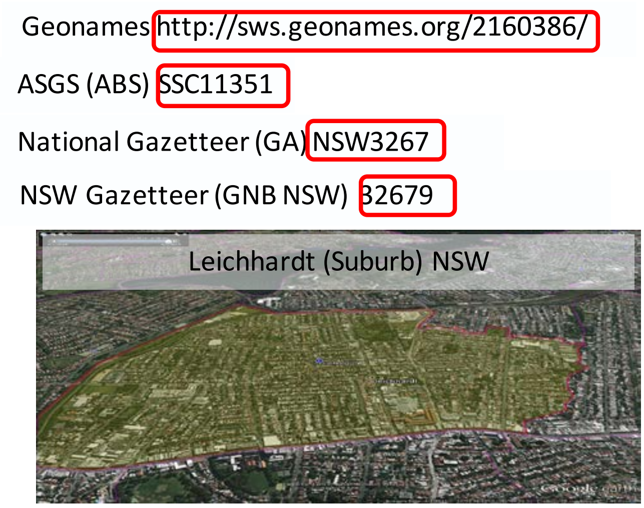

Figure 2 Multiple identifiers for the same real world object

In many cases, existing supply chains are inefficient as they rely heavily on human users to interpret and transform data structures while moving between Geographic Information Systems (GIS) environments and data formats. Extract Transform Load (ETL) tools (such as Safe software's FME) are heavily used. However, these rely on data models to enable mapping between the supplied data and target data models so that data transformations can be performed.

During these processes, errors can be made in interpretation and human processing, and meaning can be lost in translation. Furthermore, supply chains are often ungoverned with suppliers able to arbitrarily change structure, format, frequency of delivery, condition of access and use. When updates to supplied data are provided, content is expected to change. However, in many cases, structure, semantics and naming, including identifiers for features, could also be changed, arbitrarily requiring the mapping process to be repeated.

1.3.3 KEY CHALLENGES

To achieve the desired outcomes, the FSDF initiative must address a range of interwoven technical and social challenges caused by the fragmented and heterogeneous production, management, supply and governance of geospatial data across multiple levels of government.

The most critical challenge to be addressed across most of the FSDF themes is the need to integrate a patchwork of data sources with different structures and semantics (or meaning), developed under different business contexts, into a coherent suite of maintainable national products. This challenge is largely a function of the federated government structures in which spatial data production and delivery takes place across all levels of government. Foundation data products are often merely a by product of protocols related to local regulatory or business activities. INSPIRE, a pan-European system of systems, is addressing a similar challenge - that of integrating member states' data to create a seamless European Union (EU) SDI.

Important dimensions of this challenge for the FSDF are:

- **Demand-driven products** – for foundation data to serve a wider range of functions, design of products must be demand-driven to explicitly meet current and future end user needs.
- **Optimising supply chains** – reconciling heterogeneous information management and delivery frameworks across levels of governments in Australia and New Zealand to increase efficiency and address data quality issues at various stages in the information supply chain; and
- **Enabling policy settings** – addressing the heterogeneous and sometimes incompatible legislation, policies, licencing, governance and access arrangements to achieve open and no-cost data outcomes.

Many of these challenges are socio-technical in nature and require changes to work practices, collaboration arrangements, and organisational culture to support the adoption of technical solutions, through for example, processes, methods and conformance initiatives. Analogous issues and challenges exist in the broader (non-geospatial) information industry that can provide valuable lessons to the FSDF.

1.3.4 DRIVERS FOR CHANGE

The open data agenda has provided significant drivers for increasing access to geospatial data. However, this is insufficient to address the challenge faced by the geospatial community related to the coherence of the geospatial data ecosystem, as the onus is on the user community to wrangle data for use.

In practice, most integration is done by the user, privately, with little attention to reproducibility or provenance recording. This approach can be characterised as “point to point solutions” that are brittle and do not scale. However, there is an increasing recognition of the need to address this challenge holistically at source and at scale. This is one of the key drivers of the FSDF. Section 2 of this report provides a more detailed analysis of spatial data supply chains in the context of the FSDF.

In addition, many organisations recognise the significant challenges they face in effective management, delivery and use of geospatial data. These include:

- the demands on government organisations to be more responsive, outward facing, and service oriented to provide improved services to government, industry and citizens;
- the need to meet high expectations for timely, ubiquitous data delivered across multiple online platforms;
- use of emerging information and technology paradigms for online data delivery;
- handling increasing volumes and variety of data, including remote sensed topographic mapping data, sensor data, and crowd sourced geospatial data;
- challenges to the use of authoritative data posed by data provided by industry and ‘the crowd’; and
- achieving efficiency in government.

The desire to address these challenges is a strong driver for change. More broadly, there are arguments for government to lead by example and implement Information and communications technology (ICT) reforms that enable the digital economy.

1.3.5 BLOCKERS FOR CHANGE

Some significant blockers exist that militate against change. These include:

Slowly evolving ‘installed base’ – There is a significant ‘installed base’ of information systems, technology, business processes and practices and institutional and personal relationships. This installed base is held together through standards and evolves slowly as both the standards and the communities’ adoption of them are slow.

Cost of change – The cost of changing the incumbent systems, processes and skills of people in the current supply chains is high. An example of a successful standardisation is the adoption of the Water Data Transfer Format (WDTF), where specific funds were made available to data providers to upgrade existing systems.

Infrastructure externalities – the most efficient supply chain patterns shift much of the cost to the data providers. However, data providers may benefit least from the changes unless they are users of data from other providers and thus there is little incentive to bear this additional cost. This is an example of the infrastructure dilemma in which the potential collective benefit to the community is subverted by incumbents with more narrow interests.

Risk aversion – The approach proposed here has risks and some individuals and organisations prefer the status quo rather than adopting new and perhaps more risky methods.

Lack of skills – The model driven approach to creating product suites requires skills that are in limited supply. While modelling skills exist in the wider context of the information technology industry the individuals who possess these skill typically have limited understanding of geospatial concerns. People working on the geospatial side typically have little exposure to modelling information systems.

Understanding the value proposition – The value proposition, while clear to those individuals who have had the time to study the problem and possible solutions, tends to remain opaque to people working in the current value chains that have not had this opportunity, or are not required to work across the extent of the value chain.

Identifying with the problem – Many individuals and organisation are not able to identify with the problem because they work in a limited context and cannot see the broader picture across the supply or value chain.

1.4 Geospatial data interoperability

1.4.1 LEVELS OF INTEROPERABILITY

For many applications, users need to be able to integrate spatial data from multiple sources. For example, different sources of data relating to the same object, such as two road data sets that cover two adjacent areas, or disparate data sets that characterise different phenomena e.g. hydrology and land cover. However, data integration and use is problematic as spatial data is produced for different purposes at different scales, using different methods at different times, by different organisations using different systems, and is typically delivered in different formats.

To address these challenges and facilitate user access to and use of spatial data, interoperability is required. Interoperability is 'the ability to transfer and use information in a uniform and efficient manner across multiple organisations and information technology systems' (Australian Government Information Management Office (AGIMO) 2006 p. 3).

Interoperability touches on social as well as technical concerns as institutions and communities need to participate in the development of standards and agree to adopt them.

From a technical perspective, the quest for interoperability can be characterised as a series of levels each of which address a set of concerns to achieve increasing levels of interoperation as shown in Figure 3 (Brodaric and Gahegan 2006). The critical levels of interoperability are:

- **technical interoperability** achieved through the use of communication protocols such as HTTP;
- **syntactic interoperability** achieved through the use of common data formats such as XML;
- **schematic interoperability** achieved through the use of common information exchange models; and
- **semantic interoperability** achieved through the use of common vocabularies.

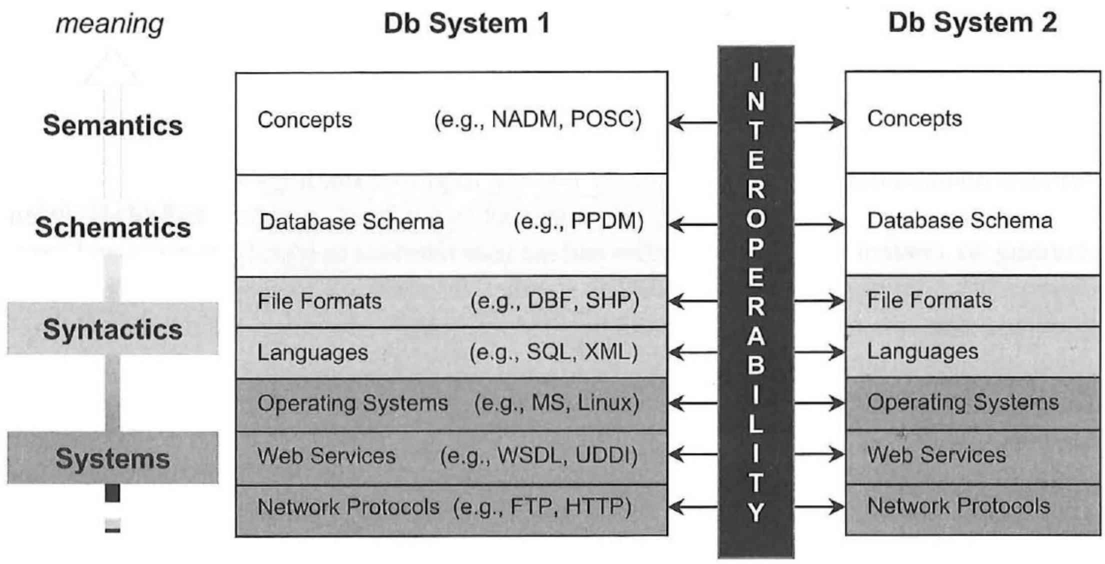

Figure 3 Interoperability levels increasing vertically for greater exchange of meaning between database systems (Brodaric and Gahegan 2006)

1.4.2 DATA FORMAT – SYSTEMS AND SYNTACTICS

At this level of interoperability, data users need to know the data delivery mechanism a data provider is using, in order to frame an appropriate data query. A wide variety of Geographic Information System (GIS) and associated technologies are used to produce, manage, deliver and work with geospatial data. These systems use a range of operating systems, proprietary data formats and communications protocols. Moving data between systems and integrating data from different systems requires using standard data exchange formats that are either file based or Web service based. Some examples include proprietary formats such as Shapefile or open formats such as Geography Markup Language (GML) (ISO 2007), GeoJSON, TopoJSON and GeoPackage. These are well-established, and many of the GIS applications can make use of them.

1.4.3 DATA STRUCTURE OR SCHEMATICS

For users that need to integrate and use multiple sources of data relating to the same spatial object, having the data sets in the same format is not sufficient. Users also need to be able to query and interpret data from multiple providers in a consistent manner and thus data needs to be available in a common structure, or schema. The need to aggregate and integrate multiple data sets is a common requirement and is in fact the key challenge that the FSDF is addressing.

However, much of the GIS software provides in-built translation tools, although some information cannot survive transformation (e.g. topology will be lost when transforming to a format that does not support topology).Common formats (described above) can be used as the basis for transmission of data and enable providers to more reliably translate from proprietary formats to a common structures, optimised for geospatial data. These provide common structures within the constraints of the underlying system, however, the configuration of data exchanged, in terms of their schematic structure or content semantics is not specified.

This issue has been previously described in ISO 19101 and ISO 19109, in the OGC Reference Model, and in the INSPIRE Generic Conceptual Model. As depicted in Figure 4, a key principle is that each ‘domain of discourse’ has a set of features and properties which members of the community talk about. These may be formalised in ‘application schemas’, each of which specifies names and data models for the feature- and property-types. Harmonisation of data from different sources confronts the basic challenge that, even if systems and syntax are agreed on, different applications result in different feature types, even for describing the same real world things. For example, communities interested in the same piece of landscape for the

purposes of mining, civil engineering, town planning, or defence, will each have a different model with different feature types.

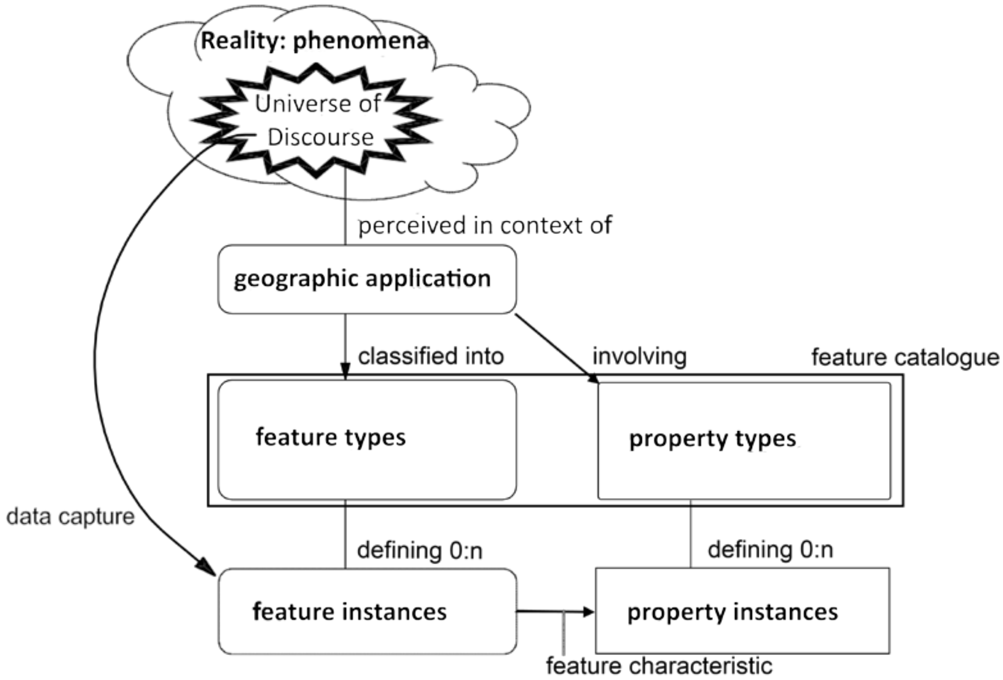

Figure 4 Modelling and representing the 'universe of discourse' (ISO 2005)

As depicted in Figure 4, geospatial data is an abstraction of reality, being (digital) *representations* of features in the world. Representations are created for various purposes, with different views of the same object, with different properties or aspects of the feature described, and often at different scales. For example, a local government area feature represented in a gazetteer (place names) data set will contain different information compared with the same feature represented in an administrative boundary data set. Similarly, a topography and hydrology data sets will contain different information about the same rivers. Each representation of the same object, even in the same system, is likely to have a different identifier, which creates problems when attempting to link information to the representations of spatial objects.

Furthermore, even in data prepared for applications of similar scope where the same view is required, the way in which the data is structured and the attributes and classification schemes used to characterise a spatial object are typically implemented in different ways in different organisations. The variations are the result of design choices made by the database or data product designers, based on interpretation of user needs, best practice, personal preference and the technology platforms being used.

The development of agreed schemas for data products is in fact the primary focus of this report with section 2, providing a description of the ways in which schemas are used in geospatial supply chains and 3, describing the methodology for developing agreed schemas through information modelling.

1.4.4 DATA SEMANTICS

Finally, data from different domains (FSDF themes) that is supposed to fit together (e.g. administrative boundaries aligned with hydrology) frequently does not due to semantic interoperability challenges. This is because different terms are used for the same concepts, different geometry, scales, generalisations, and other assumptions are used in the data preparation process, or simply that datasets originally derived from each other drift apart with time if there is no active synchronisation plan. The fact that the same spatial object may be classified in different ways in different dataset creates a semantic interoperability challenge when attempting to integrate and use two or more dataset. For example, two road datasets may use different

classification schemas to characterise road type. The issue of multiple representations and identifiers for the same spatial object highlighted in the previous section also represents semantic interoperability challenges in which users need to be able to identify which representation is actually required and to be able to integrate data that is referenced using two different identifiers.

All of this semantic variability causes confusion and imposes a cost for end users wishing to integrate and use multiple datasets representing the same object. Although data can be brought into the same environment using a common data format, the datasets cannot be queried or processed together without further interpretation, and transformation to develop concepts and classifications that are compatible. Therefore, in order to achieve interoperability of data sets that can be reliably integrated and queried in a consistent manner, there is a need to standardise the semantics i.e. the vocabularies used, within data.

1.4.5 THE ROLE OF STANDARDS

A standard is 'a documented agreement between providers and consumers, established by consensus, that provides rules, guidelines, or characteristics ensuring materials, products, and services are fit for purpose'. (OGC, ISO TC211 et al. 2014 p. 5). A key feature of standards is that new participants, on either side of the agreement, who are different from the originators, can use them.

Standards play a key role in achieving interoperability and operate at a number of levels from:

- systems– standardised communication protocols such as http and standardised delivery mechanisms such as Web feature Service or Web Map Services;
- syntactics - data formats – standard formats such as XML or JavaScript Object Notation (JSON) or patterns for using the standard formats for geographic data, such as GML, KML, GeoJSON, TopoJSON;
- schematic – data structure – community application schemas such as GeoSciML[5] and the INSPIRE theme specifications that define 'types' that appear in data; and
- semantic – data meaning – standardised vocabularies such as soil-types or units-of-measure provide values used in slots provided by the model.

While the schematic and semantic levels standards may be maintained separately and may be governed by different organisations, both are required to achieve interoperability when integrating data from multiple sources.

The adoption of standards by data providers is driven by a desire to enable the exchange, integration and use of data from multiple sources with a minimum of effort by users. Hence, geospatial standards play a vital role in enabling the development of SDI.

Two key international organisations that develop standards for geospatial information are[6]:

- The International Organisation for Standardization (ISO) Technical Committee 211 for Geographic Information/Geomatics [7]
- The Open Geospatial Consortium (OGC)[8].

ISO Technical Committee (ISO/TC) 211 has developed a series of international standards that provide a conceptual modelling framework for geospatial information. This includes constructs that define how specific aspects of spatial information should be modelled regardless of application. For example, 'ISO 19107 – The Geographic Information – Spatial Schema' (ISO 19107:2003) specifies how to describe the spatial characteristics of geographic features. These standards provide a framework within which information models can be developed for different application domains in a consistent manner. For example, the INSPIRE

[5] http://www.geosciml.org/

[6] The International Hydrographic Organization is a third international standards body concerned with hydrographic standards.

[7] http://www.isotc211.org/

[8] http://www.opengeospatial.org/

data specifications[9] are developed using ISO standards as a framework, resulting in full interoperability between data of the same theme from different providers, as well as a more limited interoperability between data from different themes.

Data specifications, such as those developed by INSPIRE, are standards that can themselves be promulgated in a specific community. This is the case for the models that are being developed under the remit of the FSDF as they will act as standards for data product development and data exchange. Both the ISO framework standards and the application schema developed from them are 'information standards'.

The OGC complements the ISO framework with a set of technology standards, which allow different systems and services to work together through the definition of standard 'interfaces'. OGC also supports the development of some specific information standards in the form of application schemas, to meet the needs of specific domains or communities. Standards such as WaterML, GeoSciML, LandInfraGML, GroundWaterML and CityGML are defined using the ISO framework and implemented to enable the delivery of data on the Web using OGC technology standard compliant Web services.

The increasing ubiquity of spatial data has raised the need for seamless integration with other data on the Web. Efforts to clarify, formalise and harmonise spatial- and Web- standards have recently commenced, through the Spatial Data on the Web Working Group (SDWWG) that has been established as a collaboration between the OGC and The World Wide Web Consortium (W3C)[10]. SDWWG will focus on determining how spatial information can best be integrated with other data on the Web; how users can discover that different facts in different datasets relate to the same place, and set best practices. It is anticipated that this activity will yield important standards for the spatial community to deliver spatial data on the Web.

Together the ISO, OGC and increasingly W3C standards provide a set of constructs that enable data to be specified, and delivered in a standardised interoperable manner. These standards provide a framework within which data products are developed. The following sections briefly describes the role of information modelling in this process.

1.4.6 INFORMATION AND DATA MODELLING

For the purposes of this report it is worth reflecting briefly on the differences between information modelling, data modelling and the nature of ontologies.

Information modelling refers to the process used to represent concepts and relationships for a particular 'domain of discourse' (or subject area). Formal information models (i.e. those which are expressed using a formally defined modelling language such as the Unified Modeling Language (UML) (Rumbaugh, Booch et al. 2004) are used to define agreed concepts and relationships. For example, we can specify that a road has a centre-line, a pavement geometry, a classification that may vary along its length, and is connected to other roads at junctions. In addition the non-geospatial properties of features can also be described (e.g. the road name, gazettal date, pavement type, usage, etc.) allowing the integration with other non-geospatial datasets. For example, the identifier for a road can be used to reference and link road maintenance data to the road feature.

The use of the UML to represent community endorsed models of conceptual entities and their relationships creates a degree of freedom between the business concerns and their technology implementations, as UML expresses information in a form that is independent of any particular technology platform. This results in systems that are potentially more accurate, flexible and interoperable.

Data modelling can be considered a form of information modelling that is concerned with the design of the logical and physical aspects of the data persistence or streaming mechanisms (typically databases). For the

[9] See Technical Guidelines Annex I http://inspire.ec.europa.eu/index.cfm/pageid/2

[10] The W3C is an international consortium that creates Web standards and guidelines http://www.w3.org/

purposes of this report the term information modelling includes data modelling. Information modelling should be conducted as a precursor to data modelling so that structure and semantics can be agreed upon in a technology neutral representation. These agreed constructs can then be implemented in technology specific ways in data models specific for a given implementation technology e.g. as Oracle or MS SQL Server spatial databases.

Ontologies are a kind of information model, though based upon distinctively different assumptions about knowledge representation than those that underpin UML modelling. While information and data models are typically concerned with database design, and supporting assessment of completeness and validity, ontologies feature an emphasis on a set of descriptions logics which enable the inferencing of additional information by considering the assertions in the data together with the axioms (or core knowledge) of the ontology. The emergence of the Semantic Web has seen the increased use of the Web Ontology Language (OWL)[11] a W3C Standard Semantic Web language used to represent knowledge about things, the relations between them and any logical axioms.

1.5 The role of information modelling in geospatial data production

Modelling the information supporting an application is the most important element of the data specification process. Formal modelling provides significant benefits when developing interrelated spatial data products, which is discussed below. Therefore, information models are the foundation of FSDF as these models will enable the production and maintenance of suites of interoperable, standards based foundation spatial data.

ISO 19100 series standards prescribe UML as the conceptual schema language. Thus, the FSDF is developing information models using UML. This approach follows global best practice, exemplified in the European INSPIRE initiative. The development and evolution of the ten FSDF themes is supported by INSPIRE-based modelling approaches, together with CSIRO modelling tools, methods and experience gained through the development of significant national and global data standards[12].

Data specification provides an opportunity to achieve:

- quality and fitness for purpose – the opportunity to update existing or develop new products that meet end user needs in the context of emerging technology paradigms;
- interoperability – achieve interoperability between products by specifying the things that need to be same across the whole of the FSDF as well as within each theme;
- harmonisation - improve coherence between data sets within and across themes; and
- efficiency – improvement in supply chains to reduce production and use costs and address common challenges which are ubiquitous across FSDF and underpin common solutions.

Supply chains are typically complex, opaque and individual organisations involved have little or no knowledge of other parts of the chain. This leads to a tightly coupled and inflexible chain. Individual actors in the chain operate autonomously seeking to maximise their own goals. Modelling the entire chain from the development of conceptual models of the data sets to the current and future needs of the users has the potential to improve the chain and bring the benefits listed above.

1.6 Value proposition for modelling

FSDF is concerned with developing a coherent framework of foundation spatial data. In this context, the value of modelling is in enabling the evolution of existing supply chains to use more efficient approaches, including

[11] http://www.w3.org/2001/sw/wiki/OWL

[12] Australian Hydrological Geospatial Fabric (AHGF) and other modelling activities in the hydrology and geosciences domains (WaterML, GeoSciMl, SoilML) and more broadly (e.g. OGC Observation and Measurements),

increased automation, machine to machine interaction and the delivery of smarter, timelier geospatial products via Web services. Key aspects of the value proposition for modelling are explored below.

1.6.1 INTEROPERABILITY

Interoperability needs to exist at a number of different levels for the value proposition of modelling to be realised. Agreeing on common semantics within a community to capture the universe of discourse for the domain will flow down from this level of analysis to the physical schema level. It will also allow cross-domain semantics to be established.

Interoperability at the schematic level will allow application schemas to be compared and integrated reducing the cost of application development because mechanisms and design patterns can be reused. This will also reduce the cost of testing and maintenance.

Standards based modelling enables interoperability between products where for example the approach used to encode hierarchy in administrative boundaries and road networks is the same. The use of consistent temporal and spatial coordinate systems between products allows the interoperability of the products.

1.6.2 IMPROVED PRODUCTS

There are typically considerable time pressures to develop and deliver products and this usually results in requirements definition being skipped or done poorly. Taking the time to define the requirements in the form of well-articulated use cases agreed by a range of stakeholders, together with a structured approach to modelling the information to meet them, will help to ensure that the right products are delivered.

User expectations, enabling technologies and requirements for geospatial data are undergoing rapid changes. Processes and products that are not developed with this awareness will ultimately fall short of expectations. An example is the emerging demand for data that can be 'mashed-up' where the ultimate representation may have come from multiple geospatial data sources. The United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM) in its 'Future trends in geospatial information management' report (Snell and Carpenter 2013), paints a picture of significant disruptive technology and changing context for geospatial information production and use. These changes represent an opportunity to re-evaluate and refine the way in which we collect, manage, process and deliver geospatial information.

1.6.3 COHERENCE BETWEEN RELATED PRODUCTS

Although modelling is critical in development of standalone information products, the investment in modelling is fully realised when models are used to develop suites of related products. In this context, models play a role in ensuring coherence between the products through the reuse of common concepts. The Australian Hydrological Geospatial Fabric (AHGF or Geofabric)[13], a suite of related national hydrological geospatial information products based on common models is exemplifies this approach (Box and Atkinson 2011). The Geofabric products provide different views of hydrological features for different purposes e.g. catchment hierarchy product, a 'blue line' cartographic product and hydrological network product, in which the same key hydrological features can be reliably identified.

[13] http://www.bom.gov.au/water/geofabric/

1.6.4 INCREASED EFFICIENCY IN PRODUCT DESIGN

The development of an agreed set of concepts that will can be reused to create multiple information products, provides significant efficiency benefits. Firstly, it provides a set of preferred or mandated concepts for use, e.g. a limited choice of coordinate reference systems. Secondly, product developers can reuse concepts developed elsewhere rather than developing their own e.g. a transport product developer can reuse the concept of a river from a hydrology model and extend of adapt it.

The ability to reuse models as design patterns also provides significant efficiency benefits. As a community defines and models its domain, patterns may emerge that can be elevated to the FSDF level and then made available to other communities. The patterns would be defined at all levels in the modelling process from *use case patterns* down to *application component patterns*.

Apart from the efficiency gains in time and effort saved, modellers would be able to move between domains and application developers could be shared across domains because the patterns would be universally understood. This would also have the effect of allowing people working in a particular domain to have more time to focus on the unique problems associated with that domain rather than concentrating on solving generic problems.

1.6.5 DATA AUDITS AND CHANGE MANAGMENT

Modelling provides benefits for individual organisations and information communities in terms of data auditing and change management processes. For individual organisations, modelling:

- provides a means to document and govern and reuse key spatial concepts within the enterprise;
- helps data owners understand and audit their own data;
- provides an opportunity to redesign products to meet use case; and
- supports the process of identifying and planning how data and systems need to change to meet identified needs.

For information communities (domains or FSDF themes) modelling:

- provides a means to agree, govern and reuse key spatial concepts within domains;
- assists the domain to understand and audit its data;
- provides an opportunity for reuse of concepts and achievement of product interoperability within and between domains; and
- supports change processes, identifying and planning how data within the domain need to change to meet identified needs.

1.6.6 MODEL DERIVED DOCUMENTATION

Communicating and understanding data and its interrelationships is critical for anyone wishing to interpret and use the data. The problem of the disconnect between information systems and the documentation that describes them has plagued the information industry since its inception. Models are created for two audiences, namely: the computational machines that will eventually process the information and create products; and the humans who need to work with the information to design the products. Traditionally, the focus has been on creating system representations for the machines and as an afterthought documentation has been largely handcrafted for the human audience. This is time consuming, costly and results in inconsistencies between the documentation and the system representations.

With model driven design there is the opportunity to derive the documentation automatically from the models. The documentation can be generated in a number of formats, representations and levels of detail for different audiences/end-uses. When the system models are changed, generation of the documentation would be triggered automatically.

There are a number of canonical types of documentation that could be produced including:

- documentation describing the models including rationales for particular design decisions, conformance with standards and the inter-relationship between models at different levels;
- a feature type catalogue describing the features, their properties and relationships with other features;
- linked data definitions that describe the features in terms of the emerging semantic Web; and
- transformation of GML (XML) data to OWL representations.

1.7 A national Spatial Data Infrastructure (SDI)

The FSDF is focused on the delivery of foundation national spatial data products. These efforts are at the core of realising a national SDI (or Spatial Information Infrastructure - SII). SDI comprises "the technology, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data" (The White House 2002).

The Australian Government is supporting the National Map initiative[14] which provides public access to open data via a Web mapping portal. Portals such as the National Map are key components of SDI providing an ability to discover and access a wide range of geospatial data delivered using Web services.

The National Map is based on a distributed architecture using Web services for the discovery, visualisation and access to spatial data via a map portal. INSPIRE is an example of a similar architecture on a pan-European scale (see section 2.6.1 for more information about INSPIRE). INSPIRE focuses on the specification of consistent European Union (EU) wide data to support environmental applications with EU member states delivering required data using the data exchange schema that have been developed.

In Australia, the national SDI may be developed in a similar fashion, with the delivery of consistent state and territory data using service services for discovery and access through a number of portals with state/territory or national scopes. To realise this goal, modelling will be critical to ensure that data held in multiple systems using different data formats, structures and semantics can transformed into a common structure and meaning (also known as an *exchange schema*) to produce national products. Delivery of data in a nationally consistent format can occur in parallel with the delivery of heterogeneous data products for use at state and territory level.

The primary challenge facing the FSDF is the delivery of coherent national products through the transformation and integration of authoritative geospatial data from Australian Governments - at commonwealth, state and territory and local government levels. This process of data supply, transformation and delivery to end users is part of a complex interwoven ecosystem of actors, supply chains, relationships and technology components that are part of national and sub-national SDI efforts.

Geospatial data supply chains need to reconcile different information models and semantics in supplied data to produce national products. Furthermore to develop a strong foundation, FSDF products need to interoperate. The next section of this report provides an analysis of existing geospatial data supply chains and the role of information modelling in enabling the design of data products, and the transformation and integration of data sets to produce them.

[14] http://nationalmap.nicta.com.au/

2 Geospatial data supply chains

2.1 Introduction

Geospatial data is created, delivered, and used through complex supply chains. These supply chains span multiple agencies and tend to be poorly understood in their entirety. As data progresses through the supply chain it may undergo data format conversion, interpretation, data transformation, integration and harmonisation with other data to produce an end product. Information models plays a key role in specifying this process as they enable the 'to be' and 'as is' structure and semantics of data to be documented and used to aid data development, interpretation and or transformation.

Modelling is used in different ways at various stages of the supply chain. To understand the role of information models, we must place them in the context of the geospatial information supply chains. This section of the report therefore provides a description of important geospatial supply chain patterns. It describes how information modelling is used in each pattern and focuses in particular on the relative costs that are borne by data providers, intermediaries (e.g. data brokers or aggregators) and users. This provides a basis for describing the role of modelling to achieve more efficient supply chains that deliver high quality timely and interoperable foundation products. This section provides an analysis of the supply chain patterns to determine the relative costs of production and total costs associated with different numbers of data users and providers for each pattern, and reveals the hidden costs borne by users.

2.2 The patterns

Based on experiences in developing and working with geospatial data production in Australia and elsewhere, a number of key patterns can be identified. These are:

- **anarchic** (or point to point) – Direct producer and user interaction without the involvement of an intermediary;
- **centralised** - Centralised production of data by a single organisation, tightly coupled application specific;
- **aggregated** - aggregation and integration of data by a single intermediary;
- **brokered** - Centralised broker transforms data to a common form from data supplied using separate models; and
- **federated** - Federated data supply using common community models.

These are distinguished primarily by the location and timing of the transformation of supplied data from source structure and format to that of the delivered 'product', using an application schema (Figure 3). The patterns are also distinguished by the actor who bears the main cost, due to the different roles in design, integration, delivery and use of geospatial data. Three primary actors are involved in the supply chains:

- provider;
- intermediary (aggregator or broker); and
- user.

For this analysis, supply chains commence with data held by providers that is processed into geospatial data products and delivered to end users. In the characterisation of supply chains presented in this report, the focus is on production and delivery of spatial data rather than its collection; data collection is therefore excluded from the analysis. Scoping the supply chain in this way, we are able to compare the relative costs for three identified actors, as well as the aggregate cost of operations associated with each pattern.

2.2.1 PATTERN 1 – ANARCHIC / POINT TO POINT

Name	Anarchic or point to point.
Summary	Direct producer and user interaction without involvement of an intermediary.
Actors	Provider, User.
Description	Each user must find, negotiate access to, extract, load, interpret, transform and then harmonise each dataset to create a coherent product with national coverage. This was the traditional pattern for GIS users – get the dataset then inspect the attribute table, and infer the meaning from column headings in order to use it in a local environment. This assumes that the tags or column headings are comprehensible to consumers. There is no intermediary.
Schema Type	Private – developed and owned by each user or organisation.
Schema use	Multiple users - each develops their own schema and integrates supplied data into this structure.
Issues	The total cost of using a product is multiplied by the number of users. This is a very inefficient and expensive model in aggregate, though it is low cost to providers as there is no requirement to transform data from their local model. The application schema may not be explicitly documented by the user. Individual users are likely to end up with a different final end product.
Exemplar	Most data sourced directly from multiple state and territory governments.

2.2.2 PATTERN 2 – CENTRALISED

Name	Centralised
Summary	Centralised supply and production of data by a single organisation.
Actors	Provider, User
Description	A single organisation is responsible for all the data that is required to produce a national product to meet a specialised internal or external user need. This organisation is either responsible for generating or collecting all the data itself, or has private arrangements with the collectors.
Schema type	Private – developed and owned by a single provider organisation.
Schema location & use	Single provider – develops the application schema based on its identified business needs.
Issues	There may be no other source of the data within Australia, and users have to accept the structure and format provided, even if it does not conform to international standards. There is likely to be varying degrees of engagement with end users in the product design process which will determine fitness for purpose of the product. There is no intermediary.
Exemplar	Australian Bureau of Statistics - Australian Statistical Geography Standard (core structures)[15] Bureau of Meteorology - Weather and climate data.

[15] http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+%28ASGS%29

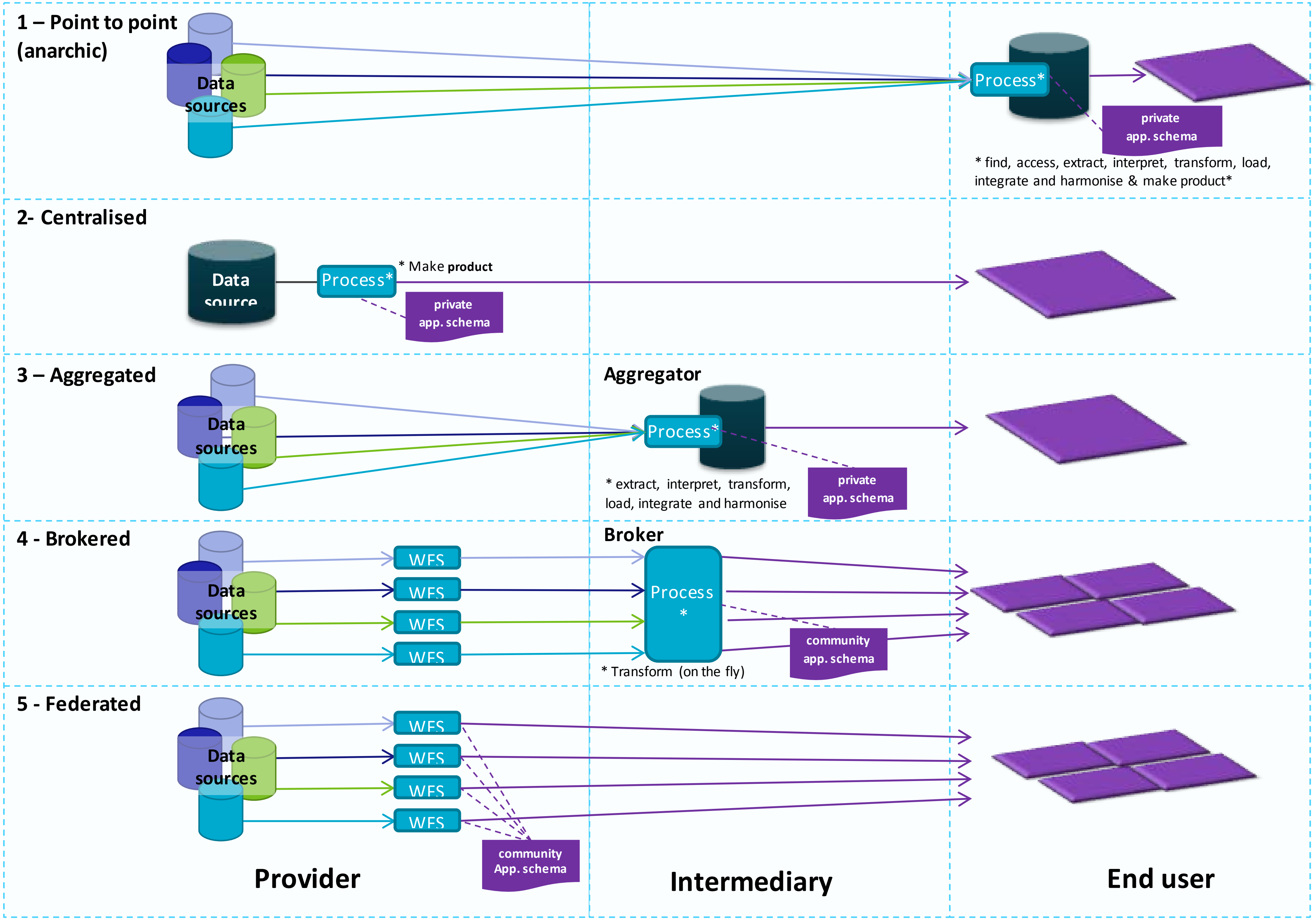

Figure 5 Geospatial information supply chain patterns

2.2.3 PATTERN 3 – AGGREGATED

Name	Aggregated
Summary	Centralised aggregation and integration of data by a single intermediary.
Actors	Provider, Intermediary, User
Description	An intermediary aggregates heterogeneous data from multiple providers, each of which publishes data according to a different structure and format, and publication method. The intermediary processes the data to create a coherent national product. Processing the data entails access, extraction, interpretation, transformation, loading, integration, harmonisation and the production and delivery of the final product. The aggregator must maintain knowledge of all the source models. The total cost of transformation scales with the number of providers. The costs is borne primarily by the aggregator, who may recover costs by charging users. The pattern represents a low cost to providers as there is no requirement to transform data from their local schema.
Schema type	Private – developed and owned by a single organisation.
Schema use	Single Intermediary - develops the application schema based on its identified business needs.
Issues	A user could obtain the data directly from each provider, and this may satisfy requirements if, for example, both providers and the application are geographically limited. However, the intermediary is the only source of a uniform national product, and can therefore charge users a premium for it in an unregulated market. A monopoly provider also has little incentive to improve quality or coverage. The size and type of user-base will be severely affected by the pricing approach.
Exemplars	PSMA Australia – G-NAF[16] Geoscience Australia - National Gazetteer[17]

2.2.4 PATTERN 4 – BROKERED

Name	Brokered
Summary	A broker transforms data supplied from services developed using a community application schema.
Actors	Provider, Intermediary, User
Description	A centralised broker service transforms heterogeneous data supplied by data providers as services in real-time to a common structure based on a community application schema. Providers use their own structure and semantics for delivery. Typically the community application schema is developed by the community but may, in limited circumstances, be developed by the broker alone. This is essentially a real-time variation on the aggregator pattern, with the cost still proportional to the number of providers, and focused on the broker.
Schema type	Community – a community agreed and owned application schema for data exchange.
Schema use	Intermediary uses the community agreed application schema to transform supplied data to a community agreed application schema.

[16] http://www.psma.com.au/?product=g-naf
[17] http://www.ga.gov.au/place-names/index.xhtml

Issues	The broker must maintain knowledge of the provider's data models. There is no additional cost to data providers, although there may be small effort required to assist the broker to map supplied data structure and semantics to the community application schema.
Exemplar	Canadian 'Groundwater Information Network' mediator[18] delivering GroundWaterML and GeoSciML EuroGEOSS broker[19]

2.2.5 PATTERN 5 – FEDERATED

Name	Federated
Summary	Federated data supply using a community agreed application schema.
Actors	Provider, Intermediary, User
Description	In this pattern the data providers provide a view of their data according to a community agreed model. Mapping of the storage data structure to the community schema is performed in a feature service hosted by each provider. End users access services from multiple data providers, with conforming to the standard (community) structure and semantics. This is helpful to users, who can use common software to process multiple sources.
Schema type	Community – a community agreed and owned application schema for data exchange.
Schema use	Multiple providers develop services that deliver data using the community application schema.
Issues	Costs in this pattern are borne primarily by the providers. Providers participate in the development of the community schemas – an upfront cost and the for ongoing maintenance of the mapping from the local schema to the community schema (arguably they are best positioned to do this). Providers are also responsible for the service performing the transformation. If the market they supply is purely external, then there may be little direct incentive or perceived payoff for the provider.
Exemplar	INSPIRE - Infrastructure for Spatial Information in the European Community[20] OneGeology[21] - Geoscience community delivering GeoSciML data

2.3 Pattern analysis

Modelling plays a critical role in improving the design and coherence of national foundation products, and in data integration. The following analysis focuses on patterns that integrate multiple sources of data to produce a national product. This is the key challenge at the heart of the FSDF.

2.3.1 CENTRALISED PATTERN

Pattern two is a prominent pattern for some specialised technical spatial data and will be the pattern used for some FSDF data. However, it is not analysed further as it does not address data transformation and integration from multiple suppliers, and thus the models are private or producer specified.

[18] http://gin.gw-info.net/service/api_ngwds:gin2/en/gin.html
[19] http://api.eurogeoss-broker.eu/docs/index.html
[20] http://inspire.ec.europa.eu/
[21] http://www.onegeology.org

2.3.2 PRIVATE VERSUS COMMUNITY APPLICATION SCHEMA

The production patterns use application schemas in different ways and at different points in the supply chain. The schema for the product is designed by the *user* in pattern one, the *provider* in pattern two and the *aggregator* in pattern three. In patterns four and five, a *community* product schema is used i.e. a schema that is agreed to by the community whose data will be transformed. The application schema in patterns one, two and three is 'private' and developed by the actor responsible for data production. This contrasts with a community schema that is developed and agreed to by a community which enables a third party to transform their data in pattern four, or enables them to supply a view of their data in pattern five. In some cases, the in broker pattern (pattern four) the schema may in fact be developed by the broker with little input from the community of providers.

2.3.3 WEB SERVICES

The broker and federated patterns are similar in so far as they deliver data using a single interface (e.g. Web Feature Services (WFS)) and both use community models to describe the conceptual aspects of the datasets. They differ in that in the broker pattern, providers deliver data using their own schema and semantics and the broker takes on the task of transforming data to a common schema, whereas in the federated pattern the providers provide a view of their data using the agreed community model.

2.3.4 RELATIVE COSTS OF PRODUCTION

In geospatial supply chains there is a total cost for data production, from data collection through to product design and delivery, to product use. For each of the patterns, the locus and total costs of production, and use of data is different. Critically, the proportion of cost borne by each actor in the chain varies greatly in each supply chain pattern. Figure 6 provides an estimation of relative costs of production (product design, development and delivery) borne by each of the three supply chain actors for each pattern[22]. As limited analysis of the cost of geospatial supply chains has been undertaken, it is not possible to present an accurate total cost for each supply chain pattern. However, there is sufficient understanding of the cost elements (design, production and delivery) of geospatial data that it is possible to depict the relative costs borne by each actor.

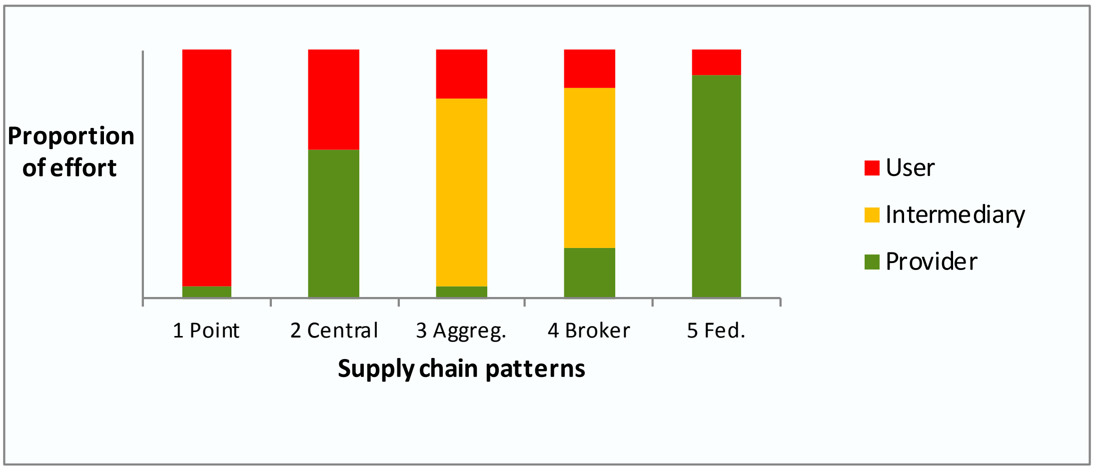

Figure 6 Relative costs of data production for stakeholder by pattern type

[22] Costs of data collection are excluded from this analysis which only compares the costs of design, production and delivery of spatial products.

As shown in Figure 6, in pattern one (anarchic) the costs are borne almost entirely by the user as the user must integrate data from multiple providers, each of which supplies its data 'as is'. In pattern two (centralised), the provider and user share the costs. The proportion of cost borne by each is determined by the nature of the product design process in particular the quality of engagement with end users to understand and ultimately meet their needs. Section 2.4 below deals with this issue in more detail.

In patterns three and four (aggregated and brokered) intermediaries bear the majority of the cost as they physically integrate and transform the data in the former or transform the data 'on the fly' according to a community schema in the latter. Providers typically participate in the development of the community application schema so they bear a small cost. In pattern five (federated) the providers bear the majority of the cost of developing the agreed community schema and of establishing and maintaining Web services to deliver the data.

2.4 The hidden cost of use

The proportion of cost borne by users in each pattern is a function of the 'fitness for purpose' of the product. This in turn reflects the amount of effort invested in product design. More analysis of use cases and user requirements equates to a higher provider cost but lower user cost.

However, considering only the relative costs of production borne by each actor does not provide the whole picture. The number of providers and users must be considered so that total cost of production and use across the entire system can be determined. Figure 7 depicts the cumulative total cost of geospatial production and use for each supply chain pattern as the number of users and providers increases.

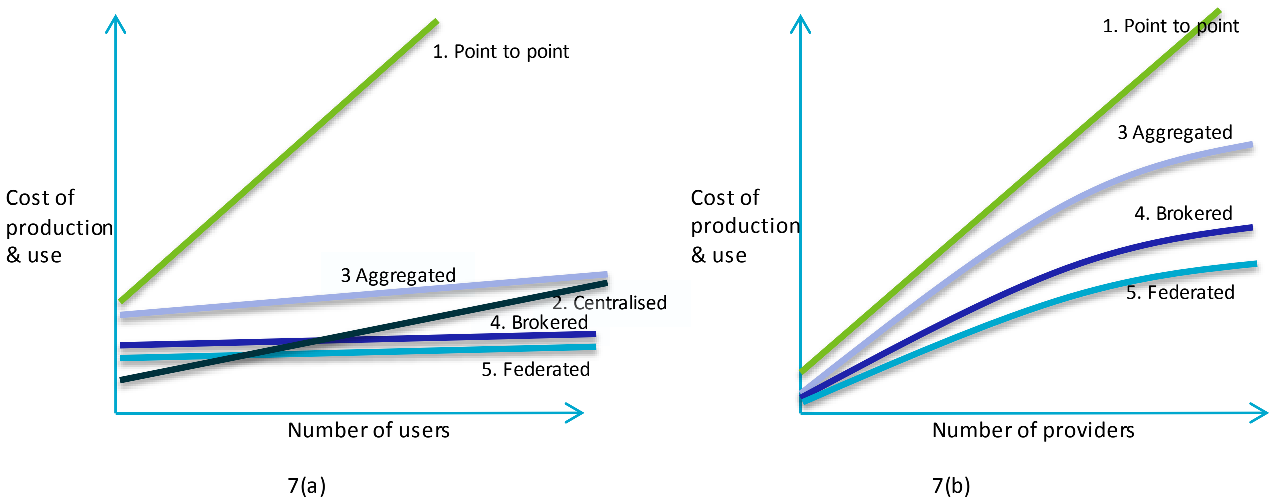

Figure 7 Cumulative cost of production and use with additional users (a) and providers (b)

Figure 7(a) shows that for each additional **user** added to the system:

- in all patterns the cost of production and use increases linearly with the number of users;
- in pattern one, the total cost of production and use increases significantly with the number of users as the cumulative cost is directly proportional to the number of users. That is each user bears the same cost of creating the product;
- in pattern two the marginal cost per additional user increases as each user will need to massage the data provided to meet their needs; and
- in patterns three, four and five the marginal cost per additional user is low. In these patterns there is a single transformation at a fixed cost per revision, so if notionally shared amongst all users, the cost per participant reduces with each additional user. However, the actual cost would be determined by the business model of the intermediary.

The steepness of the user/cost curve is inversely proportional to the effort in designing an end product to meet identified user needs. An increased investment in effort during the design phase typically results in lower user cost and effort overall as the product has a better fit for purpose. However, without any form of coordination or agreement about data, that is, a common application schema, (see pattern one user/cost curve in Figure 7), the end result is an increased cost rather than a reduction in the total cost.

Figure 7(b) shows that for each additional **provider** added to the system:

- in pattern one, the total cost of production and use increases linearly with the number of providers;
- pattern two is not applicable as it is based on a single provider;
- in patterns three, four and five there is an increased total cost for each new provider as additional effort is required to deal with each data set;
- in pattern five, costs are borne primarily by each provider so there is limited cost for users; and
- the profile of the provider cost curves for patterns three, four and five reflect the reducing marginal cost delivered through economies of scale.

It should be noted that as the number of both providers and users increases, these two factors are combined and thus totals costs for all patterns increase. However, the costs associated with the point to point pattern increase at a much higher rate than the other patterns.

In patterns four and five, the total costs of production and use are minimised by performing a transformation from provider datasets to a set of national products once. If the supplied data structure changes, a cost to remap it to a community schema is incurred. More importantly, the transformation costs are proportion to the number of input data sets as the aggregator or broker has to learn and maintain mappings from *N* data models (one for each provider) to the common model. This may be tractable for up to ten providers but becomes intractable beyond that. There are additional concerns with pattern four in that the intermediary may unilaterally decide to change the application schema or business model impacting on downstream users.

2.5 Implications for FSDF

A summary of this analysis is provided in Table 1, below.

Table 1 Supply chain pattern comparison

Pattern	Product schema	Total cost of use	Location of cost	Key characteristics
1. Point to Point	Private (User)	O(P*U)	User	Low cost for providers but large hidden cost for users
2. Centralised	Private (Provider)	O(U)	Provider	Entire supply chain owner by a single organisation. Costs borne by end users determined by investment in end user engagement by the provider
3. Aggregated	Private (Aggregator)	O(P)	Aggregator	Commonly encountered pattern. Low cost for providers but high centralised cost of production
4. Brokered	Community	O(P)	Broker	Not a common pattern. Requires investment in brokering layer for transformation
5. Federated	Community	O(P)	Provider	Onus on providers to deliver data using agreed schema

In Table 1, the total cost of production and use is the operational cost of design and delivery (O) plus the cost for each user (U) in using the data and/or the costs for each provider (P). In pattern one the number of users is multiplied by the providers to arrive at the total cost.

Supply chain patterns identified above represent a range of approaches to delivering national products. Although pattern one imposes least cost on the provider, it creates the greatest aggregate cost as every user has to do the transformation themselves. Without national efforts such as the FSDF, end users must interpret transform and integrate a number of datasets by themselves to produce a national data products.

Pattern two is encountered for some FSDF data products for example land cover, elevation and depth products. Although these supply chains may be optimal in terms of efficiency of data production, the FSDF initiative represents a significant opportunity to re-evaluate user requirements and redesign data products that are fit for purpose, given evolving requirements, delivery channels and technology platforms. Furthermore, redesign may be required so that FSDF national products are consistent and interoperate with each other.

The most commonly encountered supply chain pattern for FSDF national product development is pattern three (aggregated), with commonwealth organisations responsible for integrating state, territory and commonwealth data to produce national products. It should be noted that the aggregated pattern is also common at the state and territory level with state and territory government developing products for their jurisdictions from data provided by state or territory and local government. In some instances, this pattern may be optimal for a particular type of product, where, for example, significant processing and transformation of complex data is required. However, as for pattern two, there is significant opportunity through the FSDF to re-evaluate requirements for and potentially redesign products produced through aggregation processes. In addition, there will be opportunities to improve the efficiency of supply chains by moving from pattern three to pattern four or five.

Currently, there are no examples of the patterns four or five in operation for foundation products. Pattern four is rarely encountered and the authors believe that it is not currently in use in Australia. Pattern five has been implemented for spatial data that is not foundational (e.g. geological data).

Pattern five is the most complex from a governance perspective, as it pushes design onto the community (of data providers), and implementation onto the individual providers. It therefore imposes a cost on each provider to participate in development of the community schema and to deploy Web services to deliver data in conformance with it. However, these additional costs borne by provides are small compared to the costs associated with data use in pattern 1. Furthermore, the providers themselves will benefit through the provision of data services for local users which may reduce costs associated with servicing user requests for data. Furthermore, the data products that are compliant with community schema will be based on a deep understanding of user needs and thus may meet currently unmet user needs.

Supply chain patterns represent a continuum from uncoordinated to increasingly mature, governed approaches to delivering national products. Although patterns five is considered to be the ideal for efficient point of truth delivery of consistent state and territory data for national use, it is unlikely that all FSDF product supply chains will ultimately move to pattern five. There may be sound reasons for using patterns three or four as these may be optimal for particular types of data and/or specific circumstances. However, in some cases where there is a desire to move to federated supply chain model, financial and other factors such as technical skills shortages, policy settings, limited community cohesion or lack of demand may act as barriers for collaboration in developing community schemas and for deploying services.

It is important to reiterate that for all supply chain patterns, it is critical that data providers or intermediaries engage deeply with end users of the products to understand requirements. Ultimately the quality of products, their fitness for purpose and the reduction in the hidden cost to end users, requires investment in understanding user needs and designing products accordingly. Analysing requirements and use cases presents an invaluable opportunity to question and validate assumptions about the way in which data is used.

As can be seen from the foregoing there is a significant hidden cost in the use of products that are not fit for purpose. Thus, there is considerable potential value in assessing and redesigning products that are fit for purpose. Furthermore, there is potential to increase the efficiency of geospatial information supply chains through the development and adoption of data exchange standards by data providers. Information modelling plays a key role both in the design of products that are fit for purpose and in enabling stakeholders to transition to more efficient supply chain models to create and deliver these products. From the foregoing it posited that the benefits of investing in a data specification framework are likely to far outweigh the costs. The main benefit is found in delivering products that are fit for purpose, minimises costs of use and potentially reduced costs and inefficiency in supply chains. Broadly, the costs lie in the collective costs of implementing the data specification framework and any additional cost for each provider in deploying Web services to deliver data using an agreed schema.

2.6 International best practice and lessons learned

2.6.1 INSPIRE

INSPIRE is recognised as a world leading example of the use of model driven approaches to optimising geospatial data supply chains. While it operates at pan-European scale with EU member states, lessons learned can be applied at a national scale. A mid-term review of INSPIRE implementation has been conducted at the 5 year mark(European Commission and European Environment Agency 2014). Three key points from the review, related to data specification, are provided below.

Technical complexity

The mid-term review notes that the data specification measures put in place by INSPIRE are complex and that "no alternative could be identified in order to achieve the interoperability objective. While the actions related to interoperability are appropriate, further modifications might be taken into consideration in order to enable further benefits" (European Commission and European Environment Agency 2014)

The review recommended exploring possible modifications to the process including improved communication and opportunities to reduce the technical complexity.

Model sharing

As noted in the report "additional measures may be needed to ensure that the member states deposit and share the data models (including underlying use cases) they are detailing for individual applications. In this way they can be reused across Europe, ensuring that the interoperability achieved at the general level is not lost at the detailed one." (European Commission and European Environment Agency 2014).

Cost benefit

Estimated benefits of the INSPIRE initiative are six to seven times the cost, with estimated total investments to implement INSPIRE (including data specification and service deployment) to be between EUR 77 and 122 million per year, and estimated benefit of between EUR 770m–1 150m. A survey of practitioners undertaking Environmental Impact Assessments (EIA) and Strategic Environmental Assessments (SEAs) showed that 20,000 assessment were undertaken in Europe annually, taking an average of six months to complete, with an average of 8–10 % of the time spent finding and integrating required data. It was estimated that if INSPIRE were able to eliminate data discovery and integration costs, it would reduce the cost of conducting EIA and SEA by EUR 100–200 million per year.

2.6.2 GLOBAL EARTH OBSERVATION SYSTEM OF SYSTEMS (GEOSS) BROKER

The EuroGEOSS project has adopted a brokering approach to achieve multi-disciplinary interoperability. The brokering approach aims to lower entry barriers for both users and providers of data by taking on responsibility for transforming supplied data to a consistent structure and semantics. In this approach, data providers are not asked to implement any specific interoperability technology but to continue using their tools and publishing their resources according their standards. A brokering framework (i.e. a set of discovery access and semantic brokers) has been developed to bind heterogeneous resources published by data providers, adapting them to the tools commonly utilised by the users.

The GEOSS broker solution appears to be sustainable because of limited number of data providers. The brokering approach can also be seen as a 'bootstrapping' exercise, i.e. creating a unified supply of data to trigger development of clients. This in turn would then stimulate providers to feed the clients directly through the adoption of standards for their data, at which point the broker could be retired.

2.6.3 ONEGEOLOGY AND GEOSCIML

Rich geological data assets exist in many national and sub-national geological surveys, but are commonly difficult to discover and access. The OneGeology[23] project has been extremely successful in accessing this data by creating a federated, interoperable digital geological dataset of the planet. Its success arises from a highly coherent community with four unifying goals:

(1) to make geological data Web accessible using OGC Web Mapping Service (WMS) and Web Feature Service (WFS) standards;
(2) to transfer spatial data delivery 'know–how' to the developing world;
(3) to accelerate the development and adoption of GeoSciML[24], the international geoscience data exchange standard; and
(4) to raise the public profile and understanding of geoscience.

The OneGeology portal[25] delivers data from separate data provider services (a federated model) allowing users to access basic standardised images of maps using OGC WMS and WFS, simplified geological data using GeoSciML-Portrayal, and complex geological data using GeoSciML.

GeoSciML is a GML-based schematic and semantic geoscience exchange standard that would not have succeeded without long term commitment from OneGeology, a small number of state and national geoscience agencies and highly committed individuals. The promotion, development and implementation of federated state-of-the-art Web services and data standards such as GeoSciML by the OneGeology project serves as an exemplar for other communities.

2.6.4 NATIONAL INFORMATION EXCHANGE MODEL (NIEM)

The National Information Exchange Model (NIEM) is an XML based information exchange format for the United States (US) which has been created to facilitate information sharing across US jurisdictions. It represents a collaboration between government agencies at all levels (from Federal to tribal) and private industry. The consistent representation of common entities allows public and private agencies to efficiently share and exchange information using automated processes. The first production release of the format was in 2006 and there have been a series of revisions the latest being in 2013 testifying to the success of the endeavour.

[23] http://www.onegeology.org
[24] http://www.geosciml.org
[25] http://portal.onegeology.org/

It also worth noting that, recognising the importance of geospatial information as a key element of information sharing environments, the Geospatial for NIEM (Geo4NIEM)[26] initiative was launched. This aims to enhance NIEM's geospatial exchange capabilities and improve inter-government information sharing of geospatial data.

[26] https://www.niem.gov/technical/Pages/Geo4NIEM.aspx

3 FSDF modelling process

This section of the report provides an introduction to key modelling concepts and a description of the FSDF modelling process. The modelling approach, which places the structural definition of the information at the centre of the design process, is known as Model Driven Architecture (MDA) and has been used successfully in information technology for many years. Ideally, the information model is the only artefact that has to be maintained by the governing body.

3.1 Setting the scene

In order to set the context for the modelling process description, the following section describes key modelling concepts. The first relates to the feature-oriented view of geospatial information that is implicit in the use of ISO 19100 series and OGC standards, and underpins spatial information modelling efforts. The second describes the different types of models that are used in the data specification framework and the third provides an overview of the arrangements of models in the FSDF data specification framework.

3.1.1 A FEATURE-ORIENTED WORLD VIEW

The primary objective of modelling is to support the design and delivery of interoperable geospatial data products, primarily but not exclusively, through Web services. The OGC provides component specifications for interoperable geospatial information systems. The main focus of the specifications is on defining Web service interfaces[27] for components such as Web Feature (WFS), Web Coverage (WCS) and Web Mapping (WMS) services (see OGC Service Interfaces[28]) as part of support of service oriented architectures. However, in order to use these interfaces, it is necessary to define the model of the information that will be accessed via these interfaces. The next section provides a more in-depth description of the feature-oriented view of the world that underpins the modelling efforts. This is followed by discussion around types of models.

The General Feature Model

The OGC General Feature Model takes an object-oriented view of the world; to describe an object its type must be determined. This determines what properties are associated with it, one or more of which may be spatial.

This approach contrasts subtly but importantly with the conventional vector-GIS and Computer Aided Design (CAD) approach. In GIS and CAD, the entities of interest are characterised primarily by their geometry i.e. they are defined as points, lines or polygons; additional non-spatial attributes are added as required. Objects are often typed by the name of the layer to which they belong, which may correspond with a single feature type, though detail may also be provided by a type attribute for each geometry instance. In such conventional systems, identity is associated with geometry, and usually only one geometry (and scale) per feature is available.

An *application schema* formalises an information model for interoperability of an application domain (ISO 19109:2005), which is based on a reference model for interoperability (ISO 19101:2014). An important premise is that, since technical communication occurs primarily between members of a community operating within a universe of discourse, the language that they use is scoped to their domain, and thus should be

[27] Web service interfaces are published and standardised methods by which users can interact with web services calling specific functions of the service such as 'get map' (for WMS) or get feature (for WFS).

[28] See https://www.seegrid.csiro.au/wiki/bin/view/AppSchemas/OGCServiceInterfaces

governed by the community. Thus, the focus is firmly on a community working together to develop an agreed model of their domain. The ISO19100 suite of standards provides mechanisms for the development and implementation of geospatial tools and services. Formalisation is based on modelling the domain in terms of the feature types that occur in the domain.

Features

FSDF is primarily concerned with features, i.e. typed objects with identity such as roads, rivers or suburbs. This is often referred to as vector data in traditional GIS. Feature types are defined by a characteristic set of properties (i.e. their attributes, associations, operations). A feature type is usually specific to an application domain, and will be represented in a feature type catalogue that describes a key part of the language of a domain. Features often correspond with objects that are recognisable in the real world, but for FSDF, they relate to human constructs such as 'gazetteer', 'government unit', 'electoral boundary'. However, spatial properties are not mandatory, so a feature type could be defined for any item of interest within a domain. This potentially allows data access for both spatial and non-spatial information to be unified through a common interface.

The OGC Web Feature Service[29] is the primary interface for the delivery of feature data. Efforts in the FSDF to date have focused on modelling information for delivery via WFS.

Coverages

A coverage describes the value of a property as a function of location. The region of space-time within which a coverage is defined is called its *domain*, and the set of coverage values is its *range*. A coverage domain is often a grid of points or pixels (e.g. imagery), but may also be a network of curves (e.g. roads or streams), a tessellation of polygons, or a collection of other geometries. A coverage has values at all positions within the domain, but it is not mandatory for the domain to be continuous within its bounds.

The General Feature Model specifies that *properties* occur in the context of *features*. Hence, a coverage, which describes the variation of a property, occurs as the description of a property that varies within a feature. The OGC Web Coverage Service[30], formally defined in ISO 19128 Schema for coverage geometry and functions (ISO 19123:2005), is the primary interface for coverage data.

To date, no FSDF thematic product models have been developed for coverage data products.

Maps

A map as used in OGC refers to a portrayal of a number of (usually related) features and coverages in a single artefact. Maps are often encoded as images (e.g. GIF, JPEG, PNG) although vector portrayals are also in the same category (e.g. KML, PDF). Maps convey information to humans using colour and symbolization, and are generally not reusable for other purposes. The OGC Web Map Service[31] formally defined in ISO 19128 Web map server interface (ISO 19128:2005), is the primary interface for maps.

Coverages are distinguished from maps in that the values of the data in a coverage are on a meaningful scale, while in a map they are colours and shapes.

To date, no FSDF thematic product models have been developed for Web map products. However, WMS is typically setup in parallel with WFS to provide a portrayal service that supports the WFS data access service.

[29] See https://www.seegrid.csiro.au/wiki/bin/view/AppSchemas/OGCServiceInterfaces#Web_Feature_Service_OGC_WFS
[30] See https://www.seegrid.csiro.au/wiki/bin/view/AppSchemas/OGCServiceInterfaces#Web_Coverage_Service_OGC_WCS
[31] See https://www.seegrid.csiro.au/wiki/bin/view/AppSchemas/OGCServiceInterfaces#Web_Map_Service

Metadata records

Documents that describe datasets and services, but do not contain the actual data, are known as metadata. Such a document typically includes information related to discovery, access, and maintenance. Standards for the content and sometimes also encoding of metadata documents include Dublin Core[32], FGDC[33], ANZLIC[34], and ISO 19115 (ISO 19115:2003) and ISO 19119 (ISO 19119:2005). These documents are typically provided through an indexed catalogue interface.

The OGC Catalogue Service[35] is the primary interface for metadata records.

The FSDF thematic product models do not specify any particular metadata requirements. However, the common conceptual model does provide a placeholder for feature level metadata. In addition, the common model provides some explicit mechanisms to handle versioning of data sets, a critical aspect of metadata for products that have versions that change over time.

3.2 The FSDF model suite

3.2.1 MODEL TYPES

Traditionally, there are considered to be two main levels of model abstraction, or model types:

1. Platform Independent Models (PIM) are models that describe key concepts and relationships in a technology neutral way. These are not intended to be implemented in information systems; and
2. Platform Specific Models (PSM) - models that describe concepts and relationships in a technology specific way. These can be reverse engineered from database schemas or can be used to (forward engineer or) generate database schema.

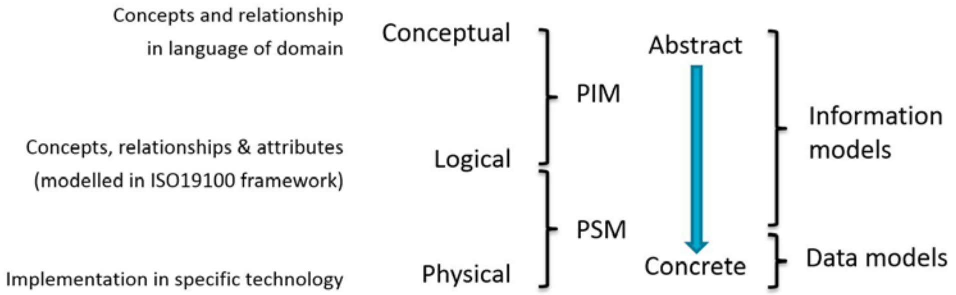

Figure 8 Modelling levels of abstraction and types of models

In practice, the boundaries between the PIM and PSM are blurred, and it is perhaps more useful to consider a continuum between purely abstract models and concrete implementations (Figure 6). Abstract, concept-only information models are platform independent, whereas database implementations, or physical models, are platform specific. However, domain specific information models, or logical models, that specify concept properties and relationships, cardinalities and property data types (such as integer, string etc.) are less abstract, but may remain platform independent. Similarly, a specific implementation of a logical model, such as a GML specification is more platform specific, but still to some extent platform independent.

[32] http://www.dublincore.org/
[33] http://www.fgdc.gov/metadata
[34] http://www.anzlic.org.au/infrastructure_metadata.html
[35] https://www.seegrid.csiro.au/wiki/bin/view/AppSchemas/OGCServiceInterfaces#Catalogue_Service_OGC_CSW

3.2.2 COMMON AND THEMATIC CONCEPTUAL MODELS

The FSDF Common Conceptual Model defines concepts and relationships together with patterns that are common across all the domains of interest i.e. the FSDF themes. It enables coherence across the domains and may include:

- the definition of the spatial and temporal representations of and between spatial objects;
- reference to common spatial and temporal reference systems as well as multilingual thesauri;
- the patterns for unique object identifiers; and
- any cross-domain constraints.

FSDF thematic conceptual models define the concepts, relationships and patterns agreed by a domain i.e. an FSDF Theme.

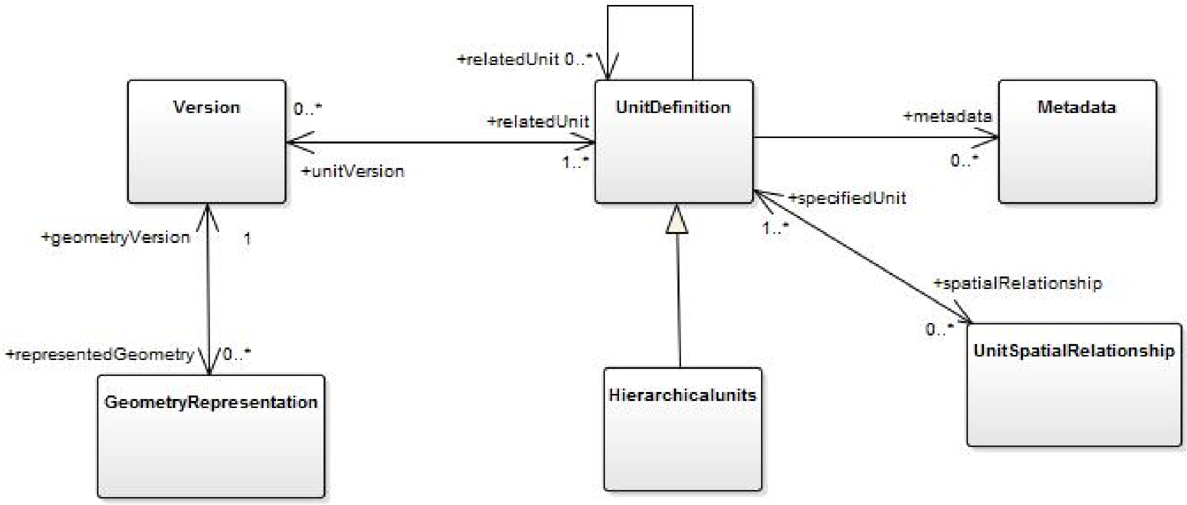

Figure 9 FSDF Common Conceptual Model

Thematic conceptual models describe the domain from an information system perspective; that is, the concepts and relationships in the FSDF theme that are to be represented in a suite of information products. This provides a conceptual model for the suite of features and properties required to meet the theme (domain) use cases. This is necessary to ensure that the suite of products within the theme can be articulated together.

At present, FSDF thematic conceptual models have not been formally documented in the UML, but instead incorporated into the design constraints of the thematic logical models. Efforts have focused on defining initial FSDF thematic logical models, and the core common logical model. A FSDF core common conceptual model as shown in Figure 9, has been developed from the logical model to document abstract concepts and their relationships including cardinalities.

It should be noted that there is likely to be considerable variation in the content that different communities agree to capture in their conceptual model. For example, with the GeoSciML community[36] (which is analogous to an FSDF theme), the UML conceptual model was purely abstract[37], the UML logical model contained significant platform specific GML components and the physical model was the XML representation of the logical model. The Observations and Measurements UML model (ISO 19156:2011) is

[36] http://www.geosciml.org/
[37] http://ngmdb.usgs.gov/www-nadm/

an example of a cross-domain conceptual model with a degree of abstraction removed, and Observations and Measurements XML (OMXML) as the physical model. FSDF thematic conceptual models are therefore likely to vary considerably in terms of the scope, nature and level of detail.

3.2.3 THEMATIC LOGICAL MODELS

The FSDF thematic logical models describe the concepts, their relationships, properties, and the data types required to meet the domain specific use cases and ensure the information products are related and remain coherent. The thematic logical models show a detailed representation of some or all of a theme's data, independent of any particular data management technology. It is the lowest common denominator that is able to encompass the stakeholder's existing physical models. It is developed by comparing the existing implementation (physical) models and is then tested against the thematic conceptual model to ensure it complies with the concepts expressed in this model.

A thematic logical model for an application domain can be formalised in an application schema using UML (ISO 19109:2005). A general methodology for the development of an application schema is provided in ISO 19109. A key premise is that communication and primary interoperability concerns centre on a community that shares a model or view of their world, and that the design of an application schema should be scoped to an information community.

The core of an application schema is a catalogue of feature types (IS19110:2005), following the General Feature Model[38] (ISO 19109:2005, ISO 19101-1:2014). The application schema is formalised[39] using UML according to the profile described in ISO 19103:2005. The FSDF application schemas use UML classes from the abstract ISO 19100 harmonised model (ISO 19103:1999, ISO 19108:2002, ISO 19109:2005, ISO 19156:2011) for standard elements, such as geometry, temporal, observations and samples.

3.2.4 PHYSICAL (PRODUCT) MODELS

The FSDF physical models describe products defined in one or more specific implementation platform(s). In this context, 'platform' may refer to the underlying technology platform used for: a data provider's database; a data product supplied by a data provider; or a national foundation data product. The FSDF thematic product models are specialisations, or profiles, of the FSDF thematic logical models, in that they constrain the features, attributes and cardinalities, of the model to meet one or more specific use cases.

The FSDF thematic product models are application schemas formalised in the GML following the rules described in ISO 19136 (ISO 19136:2007). The GML consists of a set of encoding patterns and utility components to be used in the definition of an XML document format for feature types, acting as a consistent transfer format for data in an application domain. Cartographic portrayal would however require additional styling.

The GML provides standard components for geometry, coordinate reference systems, time, measures and some other cross-domain elements. The GML specification also includes rules for construction of the application schema by extending certain base classes, in particular gml:AbstractFeatureType.

[38] https://www.seegrid.csiro.au/wiki/bin/view/AppSchemas/FeatureModel
[39] https://www.seegrid.csiro.au/wiki/AppSchemas/SchemaFormalization

3.3 The modelling process

3.3.1 OVERVIEW

The modelling process is based on the Object Management Group's Model Driven Architecture, with model design in UML using the General Feature Model from ISO 19109, the use of components from other standards in the ISO 19100 series[40], and production of the XML schema following the encoding rules specified in ISO 19136. The use of standard components for elements that are common across domains ensures maximum interoperability.

Figure 10 below, provides an overview of the modelling process, which is described in more detail in sections 3.3.2 to 3.3.6. As shown Figure 10, the modelling process is iterative with feedback loops and refinement of antecedent models based on the development and testing of models and the implementation and testing of products.

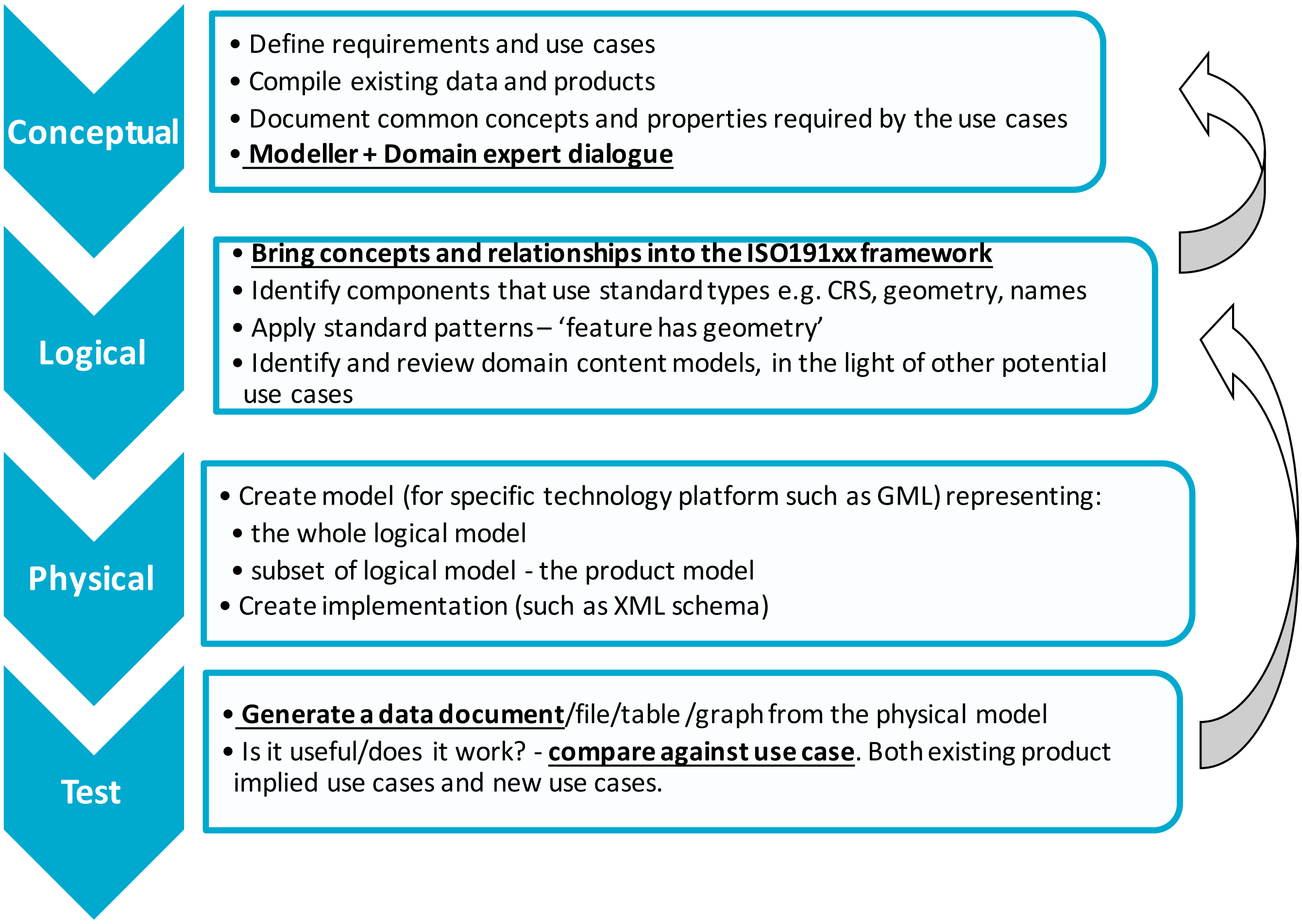

Figure 10 Overview of the FSDF modelling process

[40] https://www.seegrid.csiro.au/wiki/bin/view/AppSchemas/IsoTc211Standards

Concurrent activity

The modelling process described below, requires modelling at different meta-levels (conceptual, logical and physical) and across multiple domains; much of this process occurs in parallel. The process is iterative due to feedback from both the modelling specific to that theme, and activities associated with other themes.

The extent to which the modelling process can be followed and the level of detail will depend on the size of the community involved. For continent-wide activities, such as the European INSPIRE program, all aspects can be followed in detail. For smaller communities this may not be necessary, practical or possible. Carrying out the activity under the auspices of organisations that have established mechanisms to manage the process, such as the OGC, or concentrating on those processes important to the community may be warranted.

Key roles in modelling

The modelling process requires a variety of roles, and within each role different levels of skill are required. These include:

- domain expert;
- UML modeller;
- model reviewer;
- tester;
- service configuration specialist;
- user; and
- a number of model governance roles (see section 4.4).

Not all of these roles are engaged in the modelling process at all times, nor need they be separate individuals. Ensuring the appropriate skills are available at the right time in the process is a key to the success of the modelling process.

Modularity and change

Where multiple domains are involved in modelling activity, as is the case in the FSDF, a modular approach is required to allow each domain (or theme) to meet its own specific requirements, according to its own timing and resource capacity.

The challenge is to allow this modularity while still ensuring cross-domain interoperability and consistency. This may be achieved for large, well-funded activities, such as the INSPIRE initiative, by specifying a formal development process. This may result in as much effort being expended in specifying and implementing the processes as is required to undertake the actual modelling activities.

An alternative approach, adopted by the FSDF, is to identify a set of consistent patterns, the FSDF Common Conceptual Model, and implement these for specific domains as resources became available. This has the advantage that the modelling process is more flexible, but runs the risk that the patterns and approaches will change as each separate theme takes up the activity. Model governance and the coordination and review of models under development by an FSDF control body, will mitigate these risks. This is addressed in more detail in sections 4.4 and 5.2.2.

The modelling process should produce a model that meets identified use cases. However, with further work identifying new use cases, improved modelling patterns, or additional requirements, the models will change. Managing this change presents considerable social challenges. Model registries, versioning and governance mechanisms provide the tools to manage this change. However, assigning and carrying out the roles required to ensure appropriate change management and governance is a significant social challenge. This is especially the case where consistency is required between upgrades of loosely connected, but interdependent models.

Through the development of highly modular but interconnected models rather than large monolithic models, changes to a single model will only affect models that are directly connected to it. This is contrasted with the disruption caused by the use of large monolithic models, in which a change in one part of the model represents a new version of the entire model even though the majority of the model remain unchanged.

3.3.2 CONCEPTUAL MODEL DEVELOPMENT - REQUIREMENTS AND USE CASES

Identifying the scope and context by defining the use cases that the application schema will support is crucial to ensuring the product is useful. It is important that the scope is not too ambitious.

Generating use cases is a dialogue between the information modeller and the domain experts, both data providers and end users, carried out in a language familiar to the domain experts. The use of their language provides an introduction to the domain for the modeller.

The use cases directly determine the fitness for purpose of a data product to meet requirements. The success of the dialogue between information modeller and domain experts to generate use cases is a good indicator for the likely success and community uptake of the final information model and products developed from it.

Identifying what data products are required to meet the use cases allows user assumptions to be confirmed and new ways of solving challenges to be identified. Existing data provision needs to inform the design, but not dictate it. Although existing products reflect authentic uses cases and current best practices and community agreements, they are rarely sufficiently flexible and generic to cater for new use cases from additional data providers and users.

A significant issue concerns delimiting the community within which agreement may be reached. This may be as small as a workgroup, but may cross institutional boundaries. It may not be possible to achieve complete formal agreement amongst all interested parties, so strategies for reconciling models from different communities within the same conceptual domain will often be required.

3.3.3 CONCEPTUAL MODEL DEVELOPMENT – INFORMATION MODEL

Generating a domain specific conceptual model continues the dialogue between the information modeller and the domain experts, although now in the language of the modeller. It aims to capture the concepts and relationships required to meet the use cases, preferably in a formal language such as UML.

The differing views on what constitutes a 'conceptual model' (see section 3.2.2, above) may mean this process results in an entity that more closely resembles a logical model, i.e. feature classes that have relationship properties and attributes with cardinalities and data types specified.

3.3.4 LOGICAL MODEL DEVELOPMENT

Developing the logical model from the conceptual model entails bringing the concepts and relationships into the ISO19100 framework. The mapping between the conceptual and logical models is usually informal. That is, while both are usually UML models, no formal mapping between the two is undertaken. Rather the conceptual model is used to inform the logical model development.

Design

The General Feature Model (ISO 19109) defines feature types through their set of properties (attributes and operations). The set of properties represents the consensus of a community concerned with this feature type. The number of members of the feature type catalogue is determined by the breadth of the domain of interest, and the level of detail at which modelling takes place.

Identifying and reviewing existing domain models is important for the understanding of the breadth of information contained within the domain. Reverse engineering a provider's existing database for a community schema will typically not provide any benefit. Although it may help guide development of a community model, a specific data provider's database schema will always contain limitations when attempting to meet domain-wide requirements due to:

- the limited number of use cases the original designers considered;
- the limited range of data the schema was designed for;

- the design requirements specifically for create, read, update and delete data functionality; and
- the data provider's specific business requirements.

Additionally, more than one organisation may have a database for information which is *conceptually* the same. However, differences in the organisational requirements, or maybe just arbitrary historical reasons, may mean that the private relational database management system (RDBMS) schema is different. The "automatic" XML representations of data from these sources will therefore be different, even though the information is logically the same.

Ideally, the reasons for rejecting or modifying existing domain models should be documented.

It is important that the scope is not too ambitious and it is valuable to reuse other communities work where possible. This is particularly true for components of the ISO19100 standard suite, such as data types relating to identity, geometry, time and quantity. Patterns used in other domains should be copied, for example, 'geometry is a property of a feature, not vice versa'. They have been proven to work, and users in the domains that use the patterns will already be familiar with them.

Different parts of the model may be subject to different governance and maintenance regimes and therefore should be organised into different packages.

Packaging must take account of various considerations including:

- **stability** - Some parts will be very stable, while others will be expected to be updated more frequently; and
- **delegation** - Some parts are locally governed, while others are externally governed and adopted into the model. These must be managed carefully, and take account of the update cycle of the maintenance organisation.

In some cases, model elements belong in a package that should be under the governance of another organisation, but that organisation has not (yet) assumed responsibility, or does not provide a convenient interface. In such a circumstance, a proxy for the external package may be necessary, but it should not be entangled with the local model. Replacing the local proxy with the authoritative package should not require any other refactoring of the FSDF model.

Mutual dependencies between packages under different governance arrangements must be avoided.

The main focus of design in the model should be on the **feature types** in the application schema. In broad terms these scope the packets of information that will be transferred between systems.

Formalisation

Once the design of the model is complete and the feature types have been defined and agreed upon the effort focuses on adding detail and formalising the model. These include the following steps:

- formalise the classes (features, data types) in UML;
- apply the correct stereotypes and tagged values[41];
- specify the specialisation (inheritance) relationships;
- specify the association relationships, including their association type, navigability, role-names and cardinality;
- specify the class attributes, including their data type and cardinality;
- arrange the classes into packages where the high level packaging reflect the governance arrangements;
- enter scope notes for each:
    - Package;

[41] https://www.seegrid.csiro.au/wiki/AppSchemas/UmlGmlStereotypesAndTaggedValues

  - Class;
  - Attribute;
  - navigable association end; and
- record other rules and logic using constraints.

Hard-typing or soft-typing?

Using hard-typing (strong-typing)[42], the feature type is assigned or determined first, then the necessary properties listed and values provided. This is the usual GML encoding pattern.

When applying a full soft-typing (weak-typing) pattern, a generic feature has generic properties, and in some cases, a property that specifies (or refines) the feature type itself. For maximum flexibility, the feature has an unlimited number of properties that are soft-typed themselves. In an XML implementation this leads to less tidy instances, where semantic information occurs as element and attribute values rather than in tag names.

It is usual for a domain model to contain a mix of soft and hard-typing, determined by how much is known *a priori*, and how much flexibility the use cases require. The soft-typing approach is particularly suitable for domains where the classification of items of interest emerges from their properties, or where varying or evolving classification methodologies apply.

The use of hard- and soft-typing patterns has various advantages and disadvantages. The more detailed models, with harder typing, lead to type-proliferation and a more onerous governance and maintenance framework, suitable for a smaller community. Hard-typed data requires less effort to understand by end users. The more generic models, use softer typing with fewer types and support larger, less specialised communities with less catalogue maintenance, but require that users can cope with higher levels of abstraction.

Defining code lists is a critical aspect of the modelling process.

Logical model testing

Full testing of the logical model is usually undertaken after development of the physical model. However, a number of conformance tests can be carried out prior to this to ensure the formalisation rules have been followed.

3.3.5 PHYSICAL MODEL DEVELOPMENT

The physical model is the platform specific implementation of the logical model. For FSDF, this is a GML implementation that results in an XML schema. The physical model may either represent the entire logical model or be based on a restricted subset of the logical model to meet specific use cases. In the latter case, the product model is a profile of the logical model[43].

ISO19106 defines a profile as:

> *"Specifications of the applications of each referenced base standard or profile, stating the choice of classes or conformance subsets, and the selection of options, ranges of parameter values, for profiles;"* (ISO 19106:2004)

A feature model generated using the patterns outlined above may be converted to a GML-conformant XML schema[44]. This defines a document-model for serializing instances of geospatial data, which is suitable for

[42] https://www.seegrid.csiro.au/wiki/AppSchemas/StrongWeakTyping
[43] https://www.seegrid.csiro.au/wiki/AppSchemas/ApplicationProfiles
[44] https://www.seegrid.csiro.au/wiki/AppSchemas/GmlImplementation

transfer of data. The UML-XML encoding rules result in an XML encoding of the physical model that shows a literal and explicit relationship to the UML model.

Documentation of the physical model consists of its UML representation, its GML-conformant XML schema, and an XML conformance document describing how to configure the XML instances. This latter document details the configuration rules by specifying the requirements classes' and conformance classes, along with supporting XML instance examples. In practice, the physical model documentation is not usually completed until after the testing phase has been completed.

3.3.6 TEST THE CONCEPTUAL, LOGICAL AND PHYSICAL MODELS

Testing of the modelling process is usually restricted to two tests:

- Can the data providers' existing data be mapped to the physical model?
- Do the data documents provide the right information in the right format to meet the use cases?

Testing may also extend to establishing services delivering the data, and designing client-based filter queries. Testing focuses on data quality i.e. its fitness for purpose. However, it rarely includes performance testing.

Mapping existing data for GML applications requires generating XML instance documents that are data-rich and schema compliant.

Testing against the use cases is rarely formalised. Rather, it requires domain experts identifying whether the data being provided is likely to meet the specified use case.

3.4 Data delivery/deployment

A detailed description of the data development process is beyond the scope of this document. However this process incorporates the development of a specified data product by a provider based on the model, or the roll-out of a data exchange schema to the broader community.

As part of the data delivery and deployment process, communication within a specific information community is centred on the information model and the catalogue of feature types that is agreed to by its members. Commonly, this involves some or all of the following, depending on the size, resources and enthusiasm of the community:

- generate and publish appropriate documentation of the information models. If the formalisation process is followed then this can be auto-generated from the models;
- register the models in publicly accessible Feature type Catalogues, such as file based online registries or reusable asset services. ISO19135 provides a rigorous model for registration, in particular identifying various roles in the management of registers and the lifecycle of registered items;
- prepare and publish supporting documentation, such as the XML conformance rules and 'cookbooks';
- agree on appropriate governance arrangements;
- establish OGC WFSs using the physical model(s) at key data providers. The OGC WFS defines an interface for requesting feature data, in which the query is expressed in terms of the GML representations, and services are required to provide a GML-encoded response; and
- create client applications to consume the service data.

4 The data specification framework

4.1 Overview

The preceding sections of this report have presented the rationale for modelling, contextualised models within geospatial information supply chains and provided a description of the modelling process that has been used in the FSDF. This section describes a proposed FSDF data specification framework.

The framework is intended to support the long term development and maintenance of a suite of foundation data products. Practically, it provides the ability to:

- capture and model user requirements for data products;
- document current data products;
- unambiguously define and document concepts, relationship and classifications used within and across domains (FSDF themes);
- design and implement new products meeting articulated requirements, using agreed concepts;
- support gap analyses between current and future states for spatial products; and
- support evolution of a suite of interrelated products.

Modelling is a critical step in the data specification process, which incorporates product design and implementation. Defining the data production process, i.e. implementing products from models, is touched on only lightly in the preceding section, as this report is primarily focused on the modelling processes. FSDF models are likely to be utilised in a variety of data production contexts. These may range from an *ad hoc* process for a data product developed by a single organisation, or a more formal process defined by the FSDF to meet the needs of an FSDF theme or other defined sets of stakeholders.

The framework of models (described in section 3.2, above and section 4.2, below) is at the heart of the data specification process. It defines information standards for the FSDF. These include ISO TC211 standards (and parts thereof) used in the FSDF initiative; standardised concepts and relationships within domains (FSDF theme standards); and standardised application schema for FSDF products including data exchange schema. As such, it represents a critical component of a national SDI that could potentially be used to support communities and activities beyond the FSDF.

4.1.1 FRAMEWORK REQUIREMENTS

The following principles guided the data specification framework design:

- **foundational** – support the development and delivery of foundation spatial data as a coherent suite of interoperable products;
- **open** – provide open access to FSDF models to enable modellers and other users to access and use models to support data product development, delivery and use;
- **efficient** – deliver improved efficiency in the geospatial supply chain, by enabling modellers to discover and reuse models to develop foundation products that meet user community needs;
- **federated** – enable the governance of modular interrelated models under federated governance;
- **flexible** – provides a flexible framework enabling different approaches to the development and governance of models;
- **transparent and accountable governance** of models through their entire life cycle. Model governance should reflect and facilitate broader FSDF and spatial community governance arrangements and the federated governance of spatial information resources; and

- **effective change management** to address continual change in interdependent models and drive the development of improved products that meet new requirements and exploit emerging technology paradigms.

4.1.2 KEY COMPONENTS

As shown in Figure 11, the framework comprises three interrelated components:

1. **a suite of modular interdependent models and controlled vocabularies** that define foundational spatial data and the way in which it is used;
2. **modelling tools and processes** – the tools, systems and processes used to create, access and exploit models and vocabularies; and
3. **model and vocabulary governance** - the roles, processes, rules and mechanism for the governance of models throughout their entire lifecycle from creation, publication, use and retirement.

These components and their interactions are articulated in more detail in the following sections of the document.

4.2 The FSDF model suite

4.2.1 STRUCTURE AND HIERARCHY

The FSDF model suite comprises a hierarchy of interdependent models. Independently governed models are interrelated within the FSDF framework to achieve interoperability (based on ISO 19100 series and OGC standards) and harmonisation between foundation data within and across themes.

The models have been described in section 3.2 from a technical modelling perspective, as part of the modelling process description. The following describes the FSDF model suite from a more generic, functional perspective.

FSDF core common models

Role – these models standardise FSDF wide concerns. The FSDF common model includes dimension of data to be standardised across the FSDF, potentially including such things as spatial and temporal reference systems, together with any meta-models or design patterns that can be applied across all the themes. Core common models also include a range of context models that define aspects of the FSDF.

Structure –models are arranged in the following hierarchical package structure:

- FSDF Core Common Definition Model
    - FSDF Core Context Model – comprising models of key FSDF concepts and processes – e.g. governance structures, and modelling processes.
        - FSDF Core Conceptual Model- conceptual model of a potential geospatial information design pattern for FSDF foundation products
        - FSDF Domain Model – model key FSDF initiative concepts
        - FSDF Modelling –use case and process models that describe the data specification framework
        - FSDF Governance Model
    - FSDF Core Definition Model - logical models (application schema) of reusable design pattern and standardised aspects of FSDF data.

Current status – to date a conceptual and logical model describing important, broadly applicable design pattern for foundation spatial data has been developed. This was initially developed as part of FSDF Administrative Boundary Theme logical modelling efforts and then promoted as an FSDF core common model given its broad applicability. Discussions related to standardisation of foundation data products are yet to be commenced. Context models have also been developed to describe key aspects of the FSDF initiative.

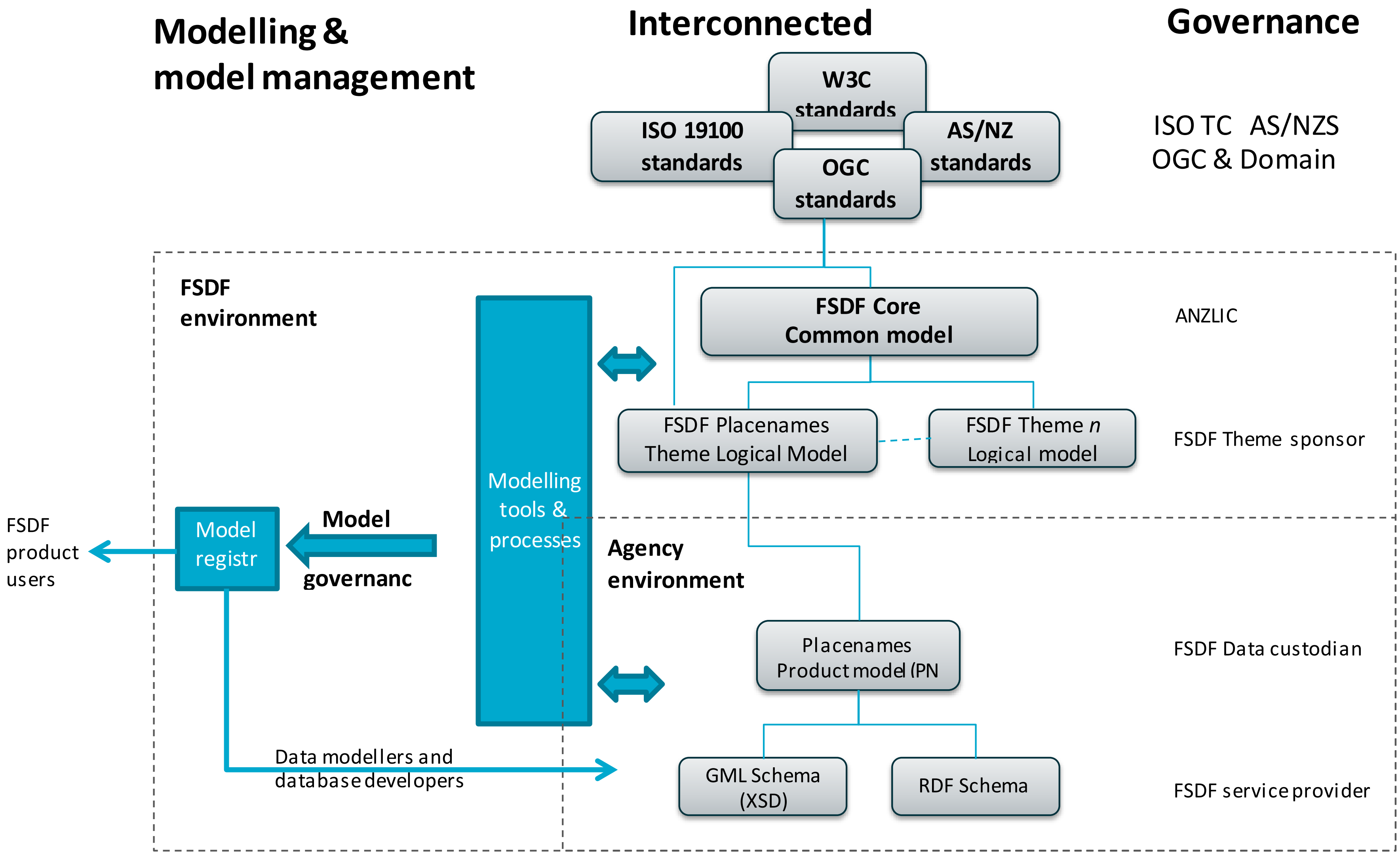

Figure 11 Overview of the FSDF model framework

FSDF thematic models

Role – defines the user requirements, use cases, key concepts and relationships for domains or FSDF themes, together with FSDF product models.

Structure – thematic models are organised in a series of packages with one package for each FSDF theme. Each theme package contains the following hierarchical package structure:

- **FSDF Requirements Model** - Use cases and other requirements that guide the development of FSDF thematic models;
- **FSDF Conceptual Model** - abstract definition of the key concepts and relationships (expressed in the language of the domain) agreed by each FSDF theme. These are unpopulated as FSDF themes have not yet engaged around conceptual model development. However, these models can include concepts that be implemented in a foundation spatial product as well as informative concepts such as business drivers, or legislative constraints that inform data product design;
- **FSDF Logical Model** - defines concepts and relationships for the FSDF theme, using the ISO19100 standards framework; and
- **FSDF Product Model** – as series of packages containing 'as is' foundation spatial data product models and designs for 'to be' implemented data products. A separate package is used for each foundation product.

Current status – To date, modelling has been focused on the FSDF Administrative Boundary and FSDF Place Names Themes[45]:

- **FSDF Administrative Boundaries** – logical models for the theme as well as a 'to be' model for an administrative boundary product have been developed.
- **FSDF Place Names Themes** - logical models for the theme as well as a national gazetteer product (as is) model has been developed.

ISO models – Hollow World

To ensure conformance with the GML standard and syntactic interoperability information modelling requires re-using packages from the ISO/TC211 ISO Harmonised Model (the ISO 19100 framework[46]). HollowWorld[47] provides a UML environment of the ISO Harmonised Model augmented with UML representations of components provided in ISO 19136 (GML). Use of HollowWorld enables specialists in a domain that utilises geospatial information to develop an information model for their application domain, which conforms to international standards for interoperability. A model strictly adhering to the modelling profile in this framework (by appropriate reuse of classes, data types, stereotypes, tagged values and other constructs) can use rule-based processing to transform it into a GML-conformant XML schema.

However, there are a number of challenges with the use of HollowWorld. The ISO 19100 suite of standards have been developed over time, with some more well-developed than others. Consequently, there are inconsistencies between the earlier and later packages. In addition, there are considerable dependencies between the packages (Coetzee, Cox et al. 2011).Of particular concern are the circular dependencies, where packages refer to other packages that in turn, either directly or indirectly, refer back to the original package. These dependencies can be strict dependencies, where an element in the package is reused, or merely referential dependencies.

[45] some initial modelling of vertical obstructions in the FSDF transport theme has also been commenced
[46] https://www.seegrid.csiro.au/wiki/AppSchemas/GeospatialStandardsContext
[47] https://www.seegrid.csiro.au/wiki/AppSchemas/HollowWorld

Navigating the ISO Harmonised Model issues requires care, and can present a considerable challenge to novice modellers. Providing the FSDF models along with only the strict dependency ISO packages via a model registry facility (see section 4.3, below) has removed some of these challenges.

4.2.2 FSDF CONTROLLED VOCABULARIES

To date, controlled vocabularies or code lists have either been defined or referenced within models or placeholders for code lists to be identified have been included in models. Emerging best practice is to reference code lists as online resources rather than actually defining them within models. This reflects the governance realities, i.e. that code lists and models may be under different governance arrangements and enables the reuse of code lists between models.

The data specification framework proposes the establishment of controlled vocabulary or code list registers for code list governance and online vocabulary services to enable access to code lists. Code lists identified or developed through modelling processes will be registered in FSDF code list registers and accessible via online vocabulary services.

4.2.3 DEPENDENCIES BETWEEN MODELS

A model is a representation of reality and is often used to allow a real world entity or concept to be reasoned about under controlled conditions. The FSDF contains a variety of information models at different levels of abstraction and across a number of different themes. These models are not isolated but rather form a fabric of articulated concepts that act as a system. The articulation is achieved by constructing relationships between the models or their parts. These relationships are best described as a series of mappings between the models allowing a set of mappings to be built that are both predictive and prescriptive.

This system does not prevent the models from being created in isolation as often is the case in the real world but it requires that the models be mapped into the system at some point in the process of their evolution, typically earlier rather than later. The intention of the models is to be a coherent and expressive representation of the knowledge and to act as a guide to ensure the correct and necessary models are created and interrelated.

The FSDF models need to be related to each other to allow traversal for a number of purposes as articulated below:

- human traversal of the model for the purpose of navigation or investigation. E.g. an architect designing a hydrological information system using FSDF data wanting to understand how a use case in a conceptual model has been realised in an implementation (product) model;
- automated traversal to allow a tool to manipulate the model. E.g. a script designed to create diagrams to show which elements in an application schema would be affected by a change to an ISO model;
- model driven transformation which utilises automated traversal to generate models of a usually lower level of abstraction. E.g. The generation of an application schema from its UML representation;
- model driven generation which utilises automated traversal to generate execution artefacts such as E.g. a GML compliant XML Schema Definition (XSD) document;
- model driven publication which utilises automated traversal to generate documentation artefacts. E.g. a data architect producing a report that defines the thematic definition of columns in database tables or the generation of a feature type catalogue; and
- resolving models from different viewpoints to create a correspondence between concepts in one viewpoint to the related concepts in another. E.g. an enterprise architect wanting to know how a concept in the enterprise viewpoint relates to a concept in the information viewpoint.

Types of mappings

The following types of mapping have been identified as useful for the relating elements and properties between models.

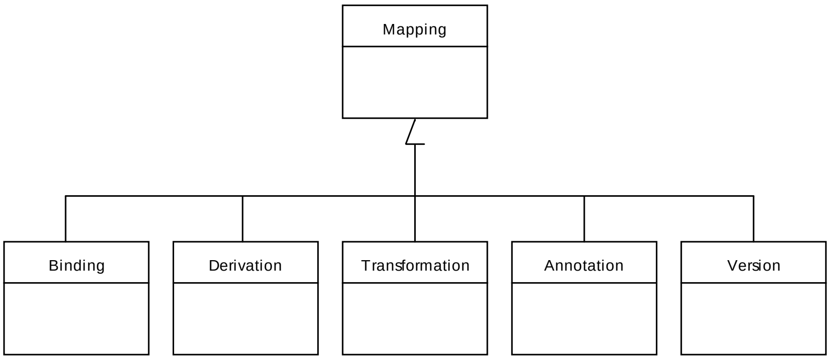

Figure 12 Types of model mapping

The following table describes each type of mapping and how it can be used. In the comparison to natural languages the models and their parts are the nouns and these relationships are the verbs of the mapping language.

Table 2 Definition of the types of mappings and their usages

Name	Description	Usage
Binding	A binding mapping relates the type of an attribute in one model to an element representing the type in another model.	Used to tighten up the definition of an attribute by constraining its instance specifications to the type definition of the bound class.
Derivation	A derivation mapping relates an element at one level of abstraction to an element at a different level of abstraction.	Used to map between different levels of abstraction to indicate the abstraction lineage of the package, element or property.
Transformation	A transformation mapping relates a package, element or property in one model to an equivalent element or property in another model typically at a different level of abstraction.	Used to relate one or more transformed elements back to the one or more pre-transformed elements for the purpose of understanding the genesis of the elements.
Annotation	An annotation mapping is a general mechanism for relating an element in one model to an element in another.	Used in circumstances where it is not appropriate to use one of the other types of mapping.
Version	A version mapping relates an element in one version in a model to an analogous element in another version of the model. The direction of the mapping is typically from a current version to a previous version.	Used to relate an element in a later version to an element in a previous version for the purposes of understanding the history of the element.

Levels of mapping

When a mapping is defined, it is essentially conveying that there is some type of relationship between the items at the endpoints of the mapping. Mappings can be performed at a number of different levels, for example, we may want to describe that two *groups* of elements have a relationship or that two elements have a relationship or that two properties of the two elements have a relationship. The properties may be

attributes, operations (services) or states. Higher level relationships should be able to be derived if a relationship exists at a more granular level; these levels are defined below.

- package (group) Level;
- element (object) level; and
- feature (property) level.

Restrictions that constrain the way these different levels can be mapped are described in the mapping grammar rules section of this document, below.

Mapping grammar rules

A natural language grammar consists of a lexicon (i.e. set of terms or words) and a set of structural rules that govern the way that the lexical items can be composed to form valid phrases and sentences. The analogous concept exists in model mapping is that the lexicon is the set of elements grouped into models and mappings, and the model grammar are the rules that prescribe how model narratives can be constructed. The following table describe the model mapping grammar rules including their rationale and application.

Table 3 Definition of model mapping grammar rules

Name	Description	Rationale	Application
Orthogonal Lines	A mapping must be horizontal or vertical, no diagonal lines allowed.	This ensures that there is that only one dimension of the matrix is traversed increasing simplicity.	Binding, Derivation, Transformation
Abstraction Adjacency	A vertical mapping must cross one and only one vertical boundary.	This ensures that there is just one level of abstraction between two mapped elements.	Derivation, Transformation
Binary Elements	A mapping must be binary meaning there must be two and only two endpoints defined.	This ensures that there is not complexities such as splitting and merging elements.	Binding, Annotation
No cycles	A series of mappings must not result in any circular references.	This ensures that a dependent model can be removed without collapsing the model ecosystem.	Binding, Derivation, Transformation
Non-Reflexive	An element cannot be mapped to itself.	This ensures that the mapping is relating two distinct elements in the model.	Binding, Derivation, Transformation, Annotation

The following matrix describes the models based on two orthogonal dimensions: level of abstraction on the vertical axis and model governance on the horizontal axis. The key phases of the process and the key models produced in these phases are shown on the y axis and the model governance level on the x axis.

Table 4 Model matrix showing the level of abstraction and model governance

		Standards	Core (Common)	Thematic
Conceptual	Requirements	Placeholder for requirements that describe the standards.	Core canonical requirements, naming and versioning requirement.	Requirements for a theme. e.g. the gazettal rules that apply to gazetteers in the FSDF Place Names Theme.
	Use Cases	Placeholders that describe the use case.	Core canonical use cases.	Use cases for a theme.

	Domain Model	Placeholders for domain model of the standards.	Core canonical domain model.	Domain model of a theme.
Logical	Logical Model	ISO logical models.	Core canonical logical model.	Theme logical model.
Physical	Application Schema Model			A GML compliant UML representation of a theme product.
	Application Schema			The GML compliant XSD for a theme product that restricts compliant instance documents.

4.3 Modelling tools and processes

4.3.1 OVERVIEW

The modelling methodology has been described in detail in section 3.3 for a technical, modelling practitioner audience. Model development and model governance are separate but interrelated concerns within the FSDF data specification framework. This section of the report provides a high level description of the modelling tools, processes and actors with section 4.4 describing the supporting model governance processes and actors.

Use cases for the specification framework are presented in Figure 13, below. Use cases include both modelling and model exploitation use cases (shown in blue) and supporting model governance use cases (shown in red). In this diagram, use cases are presented in a sequence running from top to bottom.

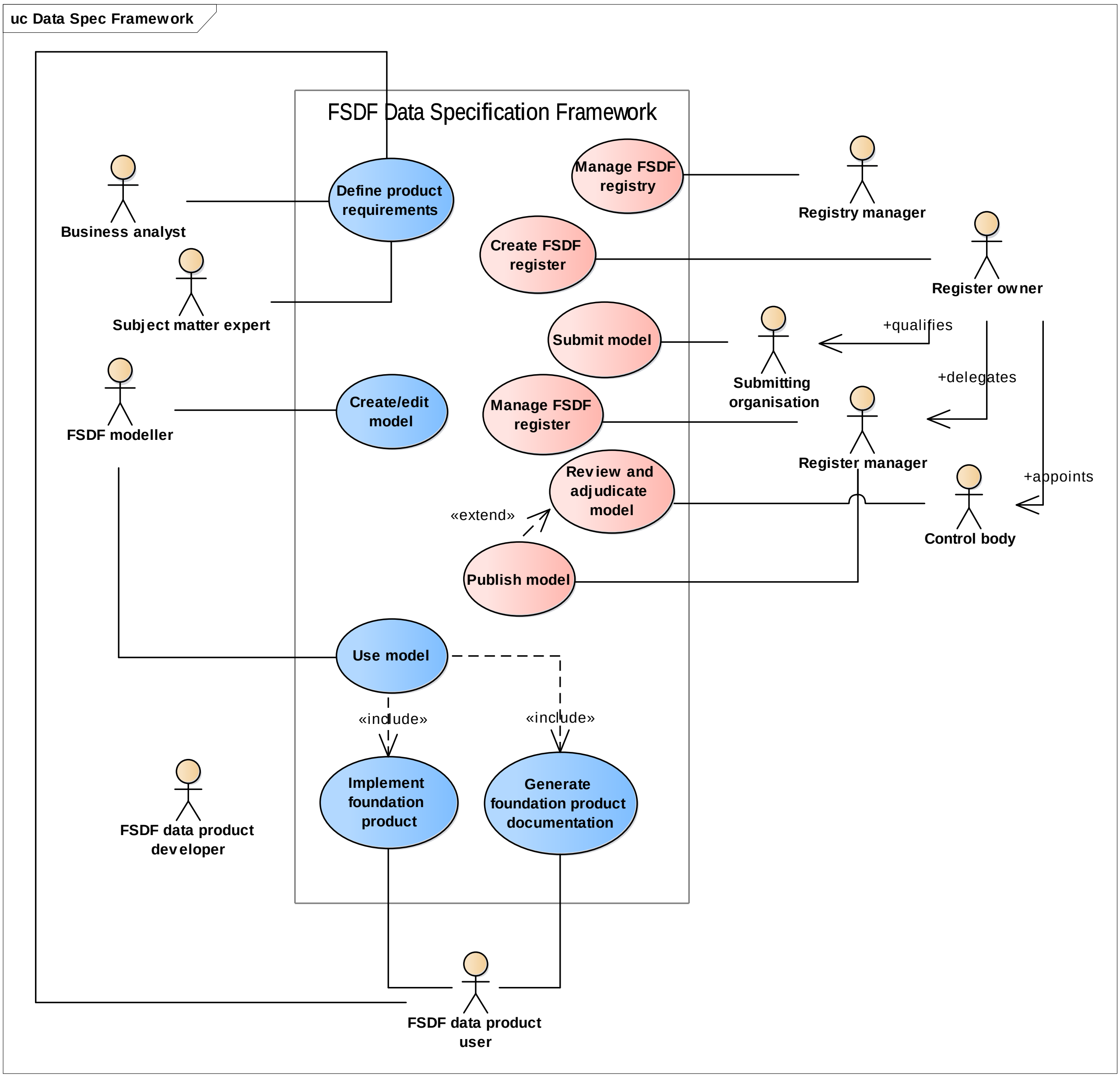

Figure 13 FSDF model framework use cases

The actors represent the key stakeholders in the modelling process and the use cases represent the goals they are trying to achieve. It is important to note that this use case model is of the modelling process and has no relationship to the use case models that are used to describe a theme's data requirements.

Modelling actors

- **FSDF modeller** – create, edit, test and (re)use FSDF models as part of the FSDF data specification process. FSDF modellers include:
  - FSDF core team modellers - responsible for developing core FSDF models; providing technical support, guidance and setting best practice for FSDF modelling efforts; providing review of FSDF models developed within FSDF themes;
  - FSDF theme modellers - responsible for developing and maintaining models within an FSDF theme;

  - FSDF product modellers - responsible for developing FSDF product models in conformance with FSDF theme models and rules and constraints specified by data custodians;
- **FSDF data product developer** – implement FSDF product models to develop foundation data products for end users. Product developers typically work within or are contracted by organisations that are custodians of FSDF products; and
- **FSDF data product user** – provide uses cases and requirements for foundation data products, and use the products and associated documentation generated from models.

Use cases

- **define product requirements** – data product users and subject matter experts define domain concepts and articulate use cases and requirements for products;
- **create and edit model** – the community of FSDF modellers create and test components of a federated, interdependent set of FSDF common, FSDF theme, and FSDF product models;
- **(re)use model** – model users access and use published models to inform development of new and the refinement of existing models. As part of this process, changes to published models can be identified and submitted to model owners. This use case includes the application schema and product documentation generation use cases;
  - **implement foundation product**– database and product developers use published models to generate an application schema for FSDF products; and
  - **generate documentation** – FSDF product users access product documentation generated directly from the model. Documentation will be provided in both human and machine readable formats and would include a feature type catalogue.

It is important to note that the model provides a single point of truth that is used to support both the development and documentation of foundation data products.

4.3.2 THE FEDERATED MODEL DEVELOPMENT ENVIRONMENT

The modelling process comprises model development and testing which is a continuous cycle. Submission of models for publication as an FSDF standard is a governance activity and covered in section 4.4, below.

Within the FSDF data specification framework, modelling occurs at three scales with differing operational and governance contexts:

- FSDF common modelling – models owned by the FSDF sponsor (ANZLIC) and developed by the core modelling team;
- FSDF theme modelling – models owned by FSDF theme sponsors[48] and produced by designated theme modellers; and
- FSDF product modelling – models owned by FSDF data custodians foundation data product owners and developed by or on behalf of FSDF data custodians.

The first two scales of modelling operate outside of individual organisational environments. According to the proposed model governance arrangements, ANZLIC would govern the FSDF environment and thus set rules and procedures for modelling. Theme sponsors would be responsible for theme modelling. Although a number of organisations act as theme sponsors, it is anticipated that a common FSDF wide approach to developing thematic models would be developed.

At present, CSIRO manages a model development environment for the FSDF common, theme and product models that have been developed. This comprises a subversion repository with version controlled model packages, and a shared modelling project in Sparx System Enterprise Architect modelling tool. Arrangements for a model development environment for common and thematic models will need to be formalised.

---

[48] At the time of writing this report Theme sponsors are Intergovernmental Committee on Surveying and Mapping (ICSM), Bureau of Meteorology, Australian Bureau of Statistics and the Department of Communications.

FSDF product modelling will typically be performed within the operational and governance context of the organisation that acts as custodian for the data being modelled. Modelling would typically need to conform to the architecture, modelling practice and standards set by the data custodian and thus prescribing this environment is not possible. However, guidelines to support the development of product models and a standardised process for validating these against thematic and common models will need to be developed. This validation of a product model would be undertaken by the relevant control body for the register to which a developed model is submitted.

4.3.3 MODELLING TOOLS

Software tools are critical to the success of the modelling effort as they have features that allow for the specification of requirements including use case and the articulation of stakeholders concerns. They simplify the creation of models and allow models to be articulated together. They also provide a mechanism for the automatic generation of human readable documentation and the creation of machine readable schemas and code. This includes the generation of GML compliant application schemas directly from the models. The ability to reverse engineer models from schema and generate schema from models for commonly used GIS systems such as ESRI using relevant data formats such as ESRI workspaces is also desirable.

Any tool that supports these features could be used. However Sparx Systems Enterprise Architect which has a cloud based registry which supports the governance of models both during their development and also in production (i.e. when they are published as standards) which makes it the tool recommended for the FSDF.

There are a number of other tools that should be evaluated and monitored as they provide useful features for information models. Of particular note, are the Humbolt Geomodel Editor[49] and the Humboldt Alignment Editor (HALE)[50] developed to support implementation of the INSPIRE data specifications.

4.3.4 MODEL REGISTRY

Models are created by a community of modellers who typically work in a distributed environment using different modelling tools. The models evolve separately over time as insights and solutions are found within domains and used to inform and refine models across domains and at more abstract levels. These models need to be carefully managed, maintained and governed to ensure they are coherent and suitable for data product specification. These models are interdependent, connected using the mappings described in section 4.2.3. A registry provides a central and managed mechanism for storing and retrieving models. These include such features as cataloguing, searching, version control. A critical function of the registry is to be ability to handle complex dependencies between multiple interrelated models that are located in existing version control repositories.

A registry is the technical mechanism that underpins the governance of the models and allows them to be managed through their entire lifecycles. It is noted that in section 2.6.1, the INSPIRE mid-term review indicated the need to be able to share and reuse models to support INSPIRE implementation. This is also a critical requirement for the FSDF given the highly modular and interdependent nature of the FSDF model suite.

An information model registry has been established for the FSDF. The Sparx Systems Reusable Asset Service[51] has been used as the software platform for the registry. The registry is hosted by Department of Finance. To

[49] http://community.esdi-humboldt.eu/projects/geomodel
[50] http://www.esdi-community.eu/projects/hale
[51] http://www.sparxsystems.com/enterprise_architect_user_guide/12.0/projects_and_teams/reuseable_asset_service.html

access the registry Sparx Enterprise Architect version 12 is required. Instructions for connecting to the registry are available on the Information Modelling page of the FSDF Website[52].

The registry currently contains, FSDF Core Common models, FSDF Admin Boundary Theme models and FSDF Placename Theme models, together with some models that describe the context of the FSDF. It is anticipated that additional registries will be established for vocabularies such as FSDF glossaries and the classification schemes, and code lists used in foundation data products/sets and the metadata that is used to describe them. An HTML view of the models developed to date has been published. This can be accessed from the information modelling page of FSDF Website[53].

4.3.5 VOCABULARY REGISTRIES

The model registry is used to register models including application schema. It provides a point of truth for information about features and attributes contained within the models. The model registry will be supplemented with a number of vocabulary registries. These are used to govern controlled vocabularies that are bound to attributes with a single register being used for each vocabulary. For example, a river feature may have the attribute called 'riverType' and a controlled vocabulary called 'riverTypeTerms' which defines the terms used to classify river types, e.g. perennial and intermittent, bound to the attribute. The vocabulary may be registered in a single register.

The governance and reuse of vocabularies in FSDF products will be critical to ensure interoperability between products. Furthermore, the reuse of vocabulary registers across information initiatives such as data.gov.au or the National Environmental Information Infrastructure[54] will assist in building interoperability between data developed and published in these initiatives, enabling it to be discovery, accessed and queried using common vocabularies.

4.4 Model and vocabulary governance

4.4.1 OVERVIEW

The FSDF has implemented a federated governance model, with ANZLIC acting as the overall authority and theme sponsors appointed with responsibility for specific themes. Within each FSDF theme, foundation data products have been identified, each of which has a custodian. Model governance will reflects these broader FSDF programme governance realities with models of different scopes being assigned to ANZLIC (FSDF common models), FSDF theme sponsors (for FSDF theme models) and FSDF product custodians (for FSDF data product models). Likewise, vocabulary governance will need to be addressed in a federated manner reflecting the realities of the ownership and custodianship of vocabularies.

4.4.2 ISO 19135 - A GOVERNANCE MODEL

Conceptually, FSDF model and vocabulary governance is based upon the ISO 19135 Standard - Procedures for Registration of Geographic Items (ISO 19135:2004). This standard articulates: the use of registers (or lists); registries (the systems that manage these lists); defined roles to establish and manage registers; and a registration process to manage the registration of items[55]. A register contains metadata about registered items such as their status, approval date and authority together with a reference to the artefact that has

[52] http://www.anzlic.gov.au/foundation_spatial_data_framework/information_modelling
[53] Ibid
[54] http://www.bom.gov.au/environment/activities/infrastructure.shtml
[55] Although ISO 19135 refers to registration of geographic item it can and has been used as model for registering a range of different resources including such things as models and vocabularies.

been registered. This metadata enables the objects being described to be managed, discovered and used to achieve common goals.

Each register is established and its contents are managed by a register owner. Optionally, the role of register manager can be delegated to another organisation. Submitting organisations (i.e. those able to submit content for inclusion in the register) are authorised by the register owner. The register owner can optionally appoint a control body to decide on submissions.

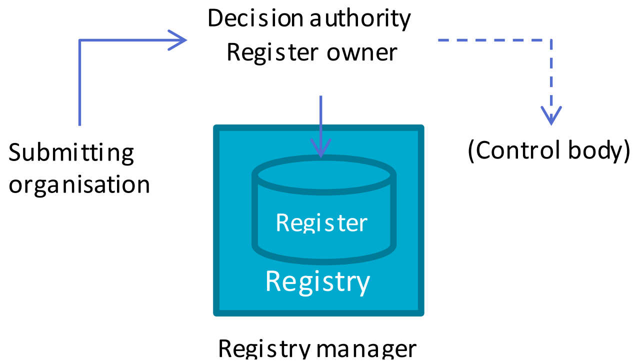

Figure 14 ISO 19135 governance roles

Repositories are typically used to store registered items. Registers can reference the location of a registered item, or can provide both a registry and repository functions, enabling access to registration metadata and the registered items themselves. At present the FSDF model registry acts as a model repository.

4.4.3 GOVERNANCE USE CASES AND ACTORS

The use case model shown in Figure 13**,** describes model governance use cases and actors. The actors are roles from ISO 19135.

- **Register owner** – a register owner is responsible for creating registers (and sub-registers) and the management, dissemination and intellectual content of those registers. This includes, determining whether submitted content should be published in the register, authorisation of submitting organisations, appointing control bodies and (optionally) delegating the register manager role to another organisation.
- **Register manager** – an organisation delegated by the register owner to manage the register on its behalf.
- **Submitting organisation** – organisations authorised by the register owner to submit models or vocabularies for inclusion in a register.
- **Control body** - optionally a register owner can appoint a control body to review submitted models and advise a register owner to accept reject or request modification of the submitted model.
- **Registry manager** – responsible for the operation of the FSDF model registry and supporting the creation of registers.

4.4.4 USE CASES

- **Manage registry** – The establishment and management of the model registry (and potentially other registries) to enable governance of and access to the models. This entails deploying, operating and administering the registry.
- **Create and manage register** - The creation and administration of registers by the register owner. It includes: establishing registers; assignment of roles and the management of permissions for

submitting organisations and optionally control bodies for its registers; creating sub-registers and delegating responsibility to register owners.

- **Submit model** – The submission of an FSDF draft model to the register for review and eventual inclusion in the FSDF framework. The model is accessible only to the register owner and the control body at this stage. Register owners are responsible for identifying and authorising the submitting organisations.
- **Review and adjudicate** – A register owner (optionally appointed control body) reviews submitted models and determines whether to accept, reject or request modification of the models. The submitted models are reviewed for conformance with ISOTC211 standards and FSDF modelling guidelines and best practice and coherence with existing FSDF models (FSDF standards).
- **Reject or request modification to a model** – A submission is rejected or a modification requested based on review.
- **Publish model** – A register owner publishes an approved model in the register as an FSDF standard, making it publically accessible.

4.4.5 FSDF REGISTRY AND REGISTER ARRANGEMENTS

Register implementation patterns will need to reflect governance reality namely, the federated nature of FSDF governance. It is proposed that registers be established for FSDF wide models, and each FSDF theme, with ANZLIC acting as FSDF register owner and FSDF theme sponsors acting as FSDF theme register owners. Product models which will be developed and owned by product custodians, would be registered in relevant FSDF theme registers.

Register owner will authorise submitting organisations and optionally control bodies to assess and advise on new items and change requests submitted to the register owner. Data product model owners would be nominated as submitting organisations for relevant FSDF theme registers.

ISO 19135 specifies three structures that can be used for registers:

- simple register – contains registered items of the same class e.g. agreements of the same type;
- multi-part register – contains registered items of different classes organised in sections based on the type of information required for each agreement type e.g. agreements of different types governed by the same authority structure; and
- hierarchical register – a structured set of registers composed of a principal register and one or more sub-registers e.g. a central vocabulary register that contains a list of the domain vocabulary registers.

A combination of all three structures is likely to be required. However, it is anticipated that the hierarchical register pattern will be commonly used, as this enables federated and partitioned governance of a particular type of registered item such as data exchange models. A hierarchical register of application schemas could for example be created with a principle register articulating the domains e.g. FSDF themes linked to domain specific sub-registers, each of which contains application schema for a specific domain.

Cross-referencing registered items between registers will be an important pattern as this will bring navigability and coherence to the agreement space. For example, a registered application schema can be cross-referenced to registered code lists used with the application schema.

5 Recommendations

5.1 Overview

This section of the report provides some recommendations for implementation of the FSDF data specification framework that was presented in the preceding section.

The primary driver for development and operation of the data specification framework is the development of application schemas that will be used to refine existing and development new FSDF national products. It should be noted that the implementation of the proposed data specification framework through the recommendations below, is considered to be worthwhile provided the framework is used to support the development of new and improved national foundation data products.

Further development of the framework should be driven by FSDF data development needs. As part of this process, suppled data products should be modelled to provide documentation of 'as is' products and to provide a basis for gap analysis between current and future states. However, further investment in developing the suite of FSDF models, implementing model governance and developing modelling tools and process, is not recommended if models serve simply as documentation of existing data products.

5.2 Recommendations

5.2.1 MODEL SUITE

Overview

The FSDF model suite has been partially developed with initial modelling of two themes (Administrative Boundaries and Place Names) and the development of a key design patterns as part of the FSDF core common model. The model suite needs to be fleshed out and mappings between models implemented.

The application schemas developed may be used in different ways depending on the geospatial data supply chain patterns used to produce a product. Thus, the application schema may be used as the basis for a new product delivered through a centralised supply chain model or may serve as a community application schema enabling the transformation of supplied data to a consistent structure and semantics using an aggregator, brokered or federated pattern.

Specific recommendations are provided below.

Core conceptual model

Recommendation 1. Articulate FSDF wide spatial data standardisation requirements and document these using the core conceptual model.
Although no FSDF wide data standardisation requirements have yet been specified, it is anticipated that as the FSDF matures, requirements will emerge. These may be identified within a theme and promoted as FSDF wide standards, or they may be developed in a top down fashion.

Recommendation 2. Test and refine the spatial object identity design pattern articulated in the core common model, through the implementation of data products.
This pattern is considered to be an important design pattern for the management and delivery of geospatial data with stable feature identifiers, a fundamental requirement for geospatial data on the Web. The pattern promotes the primacy of spatial object identity, support the explicit encoding of topological relationships

between spatial objects, feature level metadata and provides for association of multiple versions and multiple geometric representations of spatial objects to their identity.

Thematic models

Recommendation 3. Develop thematic conceptual and logical models to define the core concepts and relationships for themes.
The effort invested in this will be commensurate with the data scope and complexity of the theme and business drivers for the (re)design of national foundation products.

Product modelling

Recommendation 4. Develop use case and requirements for priority foundation products.
This provides an opportunity to revisit assumptions about what is actually required to meet end user needs without focusing on existing data delivery paradigms. Use cases developed to date are very high level, focus on the 'as is' state (rather than the future state) and provide insufficient detail to support product design. There is strong evidence, from experiences in developing the administrative boundary product, that investment in use case and requirements elicitation and analysis provides opportunity to develop innovative new products that are fit for purpose.

Recommendation 5. Develop application schema for priority FSDF data products and deploy data services. The rationale for investment in the data specification framework is the development and delivery of a suite of coherent high quality national foundation products. Therefore, further model development must be driven by data product development priorities. This may be the development of a common application schema for a Web service which suppliers commit to delivering or product application schema that will be used by a single supplier or aggregator to deliver a national product. Application schema development should be informed and constrained by the logical thematic modelling. They should also be able to accommodate existing data and thus it is important to test the application shcema using supply data. This will demonstrate how existing data can be mapped to the new schema.

Recommendation 6. Develop and implement a consistent approach to managing code lists found in foundation data. Code lists for foundation data should be managed and governed using one or more vocabulary registries with vocabulary services providing access to these code lists. This will enable code lists to be referenced and reused in a consistent manner across modelling activities enabling interoperability. An audit and registration of code lists used in foundation data is recommended. This will enable harmonisation and cross-referencing between code list and will support data harmonisation, aggregation and transformation processes.

5.2.2 GOVERNANCE ARRANGEMENTS

Interim governance arrangement have been put in place to enable the models developed to date to be published in the model registry. To implement the data specification framework, formal assignment of governance roles defined in ISO 19135 (and discussed in section 4.4) will be required. These arrangements are not onerous and can be implemented in a lightweight manner as described below. These recommendations apply both to model and vocabulary governance[56].

Recommendation 7. Appoint registry manager – an organisation should be formally appointed to act as FSDF registry manager (responsible for the ongoing operation of the registry). This organisation will be responsible for establishing and operating FSDF registries.

[56] Federated governance of vocabularies with one register per domain (FSDF them) may be required.

Recommendation 8. Determine register configuration - appropriate register structures for the FSDF should be established. This could comprise one register per theme with hierarchical sub-registers for each type of model. Alternatively, cross theme registers could be established for identified types of models e.g. an application schema register, a thematic model register etc. The latter is the recommended approach as, fewer registers will have simpler governance arrangements, and will provide users with a single point of truth for models of each type. The application schema register and the code list register will be of key importance for FSDF product implementation.

Recommendation 9. Identify FSDF register owner - a peak geospatial body such as ANZLIC or The Intergovernmental Committee on Surveying and Mapping (ICSM) should be appointed as owner of all FSDF registers. If thematic registers are established register owners for each theme register will need to be defined.

Recommendation 10. Delegate register manager – identified FSDF register owner(s) should delegate register manager(s) to manage the registers and registration process.

Recommendation 11. Appoint control body and submitting organisations the register owner should appoint a control body and identify submitting organisations for each register. It is recommended that a technical working group of organisations (possibly the existing technical advisory group or a working group within it) engaged in modelling, be formed to act as control body for registers. It is also recommended that organisations that are members of this control body should act as submitting organisations for registers. A single control body and the same submitting organisations could be used for all registers. This would assist in achieving coherence across the FSDF design and implementation.

To support ISO 19135 governance arrangements the existing model registry implementation will need to be extended to incorporate additional functionality described (see section 5.2.3 below), The governance arrangements outlined above should also be applied to controlled vocabulary (code list) governance. Recommendations for establishing controlled vocabulary registries are provided in section 5.2.3 below.

5.2.3 MODELLING TOOLS AND PROCESSES

Model development

Recommendation 12. Develop modelling guidelines – this will include further defining the processes for developing logical product models from logical thematic models and developing physical product models from logical product models. In addition, guidelines for model mapping need to be further developed and codified in modelling guidelines.

Recommendation 13. Building modelling capability - Build skills and capability for cadre of spatial information modellers through FSDF training programmes. It is anticipated that two key types of modellers will be required – expert conceptual and logical modellers, and logical and physical product modellers. It is anticipated that the development of core common models and thematic models will be undertaken by a core cadre of specialist modellers such as information architects from organisations. This cadre of modellers will also play the role of control body for model governance. The bulk of modelling activity at the product level will be undertaken by data modellers in agency environments. There is a recognised shortage of spatial information modelling skills and capability and this will need to be addressed.

Recommendation 14. Formalise cross theme model and data alignment procedures. Establish procedures for cross theme model and data harmonisation. It is recommended that this be addressed at two levels. Firstly, thematic and product modellers should be responsible for scanning FSDF models and determining potential requirements for harmonisation their models with those of other themes. This can be assisted by the analysis of existing FSDF data to identify potential overlaps of spatial objects between themes (e.g. admin boundaries and place names or transport and hydrology). Secondly, the assessment process undertaken by

control bodies when reviewing a model that has been submitted for publication, should explicitly assess the model against published models to identify any missing or problematic harmonisation aspects of the models.

Model custodianship

Governance roles above are distinct from model custodianship. Custodianship of models should be addressed in a similar fashion to that of data, with clear identification of owners for each of the models. It is worth noting that model ownership is split between community in the case of FSDF core common and thematic models and specific organisations in the case of product models, as product models are owned by custodians of the data which they describe.

Recommendation 15. Peak geospatial body (such as ANZLIC or ICSM) act as custodian of the FSDF core model.

Recommendation 16. FSDF Theme sponsors act as custodians of the Theme models. In addition theme sponsors are responsible for coordination of modelling activity within their theme.

Recommendation 17. Determine custodians for FSDF product models. FSDF data product custodians would be best placed to act as product model custodians.

Model development environment

The model registry provides a means to manage the standardisation process, enabling models to be submitted, reviewed and published as FSDF standards. Prior to submission, models need to be developed. Thus a shared modelling environment, enabling communities of modellers to collaboratively develop models, is required. Currently, CSIRO operates an enterprise scale (i.e. within CSIRO) shared modelling environment. However this is an informal arrangement and does not meet the needs of the FSDF community.

Recommendation 18. Formalise arrangements for an FSDF model development environment. Arrangements for a cross enterprise modelling environment need to be formalised. It is proposed that the Sparx Cloud Server (model registry) shared model, hosted by a designated agency and underpinned by a subversion repository be used as an FSDF modelling environment. On completion of a model in the development environment, it would be submitted to the FSDF registry for review by a control body and eventually, publication.

Model registry

A registry is the technical mechanism that underpins the governance of models and allows them to be managed through their entire lifecycle. As noted in section 2.6.1, the INSPIRE mid-term review indicated the criticality of model sharing and reuse. The current registry implementation provides a basis for model registration but additional functionality needs to be developed to support ISO19135 governance and to enhance access to models and model elements to support documentation use cases.

Recommendation 19. Implement ISO 19135 roles and registration metadata in the model registry. The current registry implementation has limited permission control. This needs to be extended to support ISO 19135 registration roles. Support for ISO 19135 registration process metadata also needs to be implemented in the registry to enable the management of registered items through the registration process.

Recommendation 20. Develop an application programming interface (API) on the registry to enable access to models and model elements. Currently, the only way to access models and model elements in the registry is through use of the Sparx Enterprise Architect tool. Once models have been registered in the registry it is highly desirable to publish them as documentation in human and machine readable forms in semantic Web formats such as Web Ontology Language (OWL)[57] and Simple Knowledge Organization System

[57] http://www.w3.org/2001/sw/wiki/OWL

(SKOS)[58]. To support the wide range of anticipated use cases, it is recommended that a programmable interface to the registry be developed.

Vocabulary registry

The model registry enables the governance of application schema and the framework of FSDF models that support their development. Application schema describe structural aspects of data, namely feature types (e.g. road, river, or administrative boundary) their properties and relationships to each other. When modelling spatial objects, the classification schemes (i.e. code lists) used to characterise properties of spatial objects also need to be governed and managed so that they can be reliably accessed and (re)used by modeller in data products modelling and by end users of to aid product interpretation and use.

Recommendation 21. Implement vocabulary registry and vocabulary services. Mechanisms for FSDF vocabulary registration and delivery should be implemented. These comprise a vocabulary registry with a vocabulary register (a list of the vocabulary lists) and sub-registers for each registered vocabulary. Two types of controlled vocabularies are anticipated - general FSDF vocabularies and code lists used in FSDF information models. It is anticipated that controlled vocabularies such as FSDF governance role vocabulary or FSDF theme register, for adoption by the FSDF, will be identified through FSDF programme management activities. Code lists to be adopted by FSDF will be identified through modelling activities and may comprise externally governed code lists or those developed and governed by the FSDF. Examples include e.g. ISO or OGC governed vocabularies (e.g. temporal terms), FSDF cross theme vocabularies (e.g. data quality terms developed for the FSDF) or theme specific terms (e.g. a national road classification scheme).

[58] http://www.w3.org/2004/02/skos/

Acronyms

ABS - Australian Bureau of Statistics

ANZLIC - The Spatial Information Council

API - Application programming interface

ASGS - Australian Statistical Geography Standard

EIA - Environmental and Strategic Impact Assessments

ETL - Extract Transform Load

EU - European Union

FSDF - Foundation Spatial Data Framework

FTP - File Transfer Protocol

GA – Geoscience Australia

GDI - Geospatial Data Infrastructure

GEOSS – Global Earth Observation System of Systems

GIS - Geographic Information Systems

GML- Geography Markup Language

HALE- Humboldt Alignment Editor

ICSM - Intergovernmental Committee on Surveying and Mapping

ICT – Information and communications technology

INSPIRE – Infrastructure for Spatial Information in the European Community

ISO – International Organisation for Standardization

LGA – Local Government Areas

NADM – North American Data Model

MDA - Model Driven Architecture

NICTA – National ICT Australia

OGC – Open Geospatial Consortium

OWL - Web Ontology Language

POSC – Petrotechnical Open Standards Consortium

PPDM – Public Petroleum Data Model Association

PSM – Platform Specific Model

PIM – Platform Independent Model

SDI – Spatial Data Infrastructure

SKOS - Simple Knowledge Organization System - home page

SEA - Strategic Environmental Assessments

UML – Unified Modelling Language

WDTF - Water Data Transfer Format

WFS – Web Feature Service

WMS – Web Mapping Service

XML – Exchange Markup Language

XSD - XML Schema Definition

Glossary

Abstraction - a conceptual process of reducing the information content of a concept or an observable phenomenon, typically to retain only information which is relevant for a particular purpose. "An abstraction" is the product of this process—a concept that acts as a super-categorical noun for all subordinate concepts, and connects any related concepts as a group, field, or category.(Wikipedia contributors 2015). See also 'level of abstraction'

Application programming interface (API) - a set of routines, protocols, and tools for building software applications. An API expresses a software component in terms of its operations, inputs, outputs, and underlying types.

Application schema - A set of conceptual schema for data required by one or more applications. Application schemas are information models for a specific information community. (Open Geospatial Consortium 2015). See also - information model.

Cardinality - the number of elements in a set. Relationships between classes in information models and between tables in a database have cardinality expressed in relation to each other e.g. one to one, one to many or many to many relationships between classes or tables.

Conceptual model – an abstract model of the concepts and relationship of a domain of discourse.

Control body – A group of technical experts that makes decisions regarding the content of a register. (ISO 2004)

Data (product) specification – A detailed description of a data set or data set series together with additional information that will enable it to be created, supplied to and used by another party. (ISO 2008)

Data specification framework – A framework of governance, processes and tools that supports the development and management of data specifications for a suite of data products.

Domain - An area of knowledge or activity. In the governance context domain refers to the extent of control of a governing authority. (Merriam-Webster 2015) (e.g. ruler, government, decision authority).

Domain of discourse - (also universe of discourse) the set of entities or objects considered within a given context (ed. Prokhorov 1983)

Feature - Abstraction of a real world phenomenon. "A digital representation of a real world entity or an abstraction of the real world. It has a spatial domain, a temporal domain, or a spatial/temporal domain as one of its attributes. (Open Geospatial Consortium 2015)

Feature Type - A digital representation of a real world entity or an abstraction of the real world. It has a spatial domain, a temporal domain, or a spatial/temporal domain as one of its attributes (Open Geospatial Consortium 2015)

Feature (Type) Catalogue - Catalogue containing definitions and descriptions of the feature types, feature attributes, and feature relationships occurring in one or more sets of geographic data, together with any feature operations that may be applied. (Open Geospatial Consortium 2015)

Federated – Independent entities (organisations, political or territorial units) joined in an alliance (federation), ceding some powers and decision authority to the federation level while retaining control and decision authority for local matters (i.e. over its own territory or organisation operations).

Geospatial - Referring to location relative to the Earth's surface. "Geospatial" is more precise in many GI contexts than "geographic," because geospatial information is often used in ways that do not involve a graphic representation, or map, of the information. (Open Geospatial Consortium 2015)

Geospatial Data Infrastructure (GDI) – see Spatial Data Infrastructure

Geography Markup Language (GML) – OGC`s XML based language for describing and encoding geospatial information. An application of XML, a specification developed by members of the Open GIS Consortium. http://www.opengis.org/techno/specs/00-029/GML.html ". GML is an XML encoding for spatial data. In a sense, it is a schema-writing language for spatial information. (Open Geospatial Consortium 2015)

Governance – A framework of 'authority structures' and processes, by which communities manage their collective affairs through a continuous process of negotiation and decision making. The framework enables the creation and operation of mechanisms, processes and rules designed to reconcile the diverse needs and interests of a community, to steer individual and collective initiatives of stakeholders to achieve agreed, collective goals. (Box 2013).

Harmonisation – is a process through which different standards can be brought together through alignment, coordination, blending or matching of components parts to create a coherent whole. Standards harmonisation is an activity undertaken by communities of experts to align standards. (Open Geospatial Consortium 2015)

Information community - A collection of people (a government agency or group of agencies, a profession, a group of researchers in the same discipline, corporate partners cooperating on a project, etc.) who, at least part of the time, share a common digital geographic information language and common spatial feature definitions. (Open Geospatial Consortium 2015)

Information Infrastructure – interconnected systems with interwoven social and technical components including information supply chains, institutional arrangements, standards, and technology. See also: Spatial Data Infrastructure, System of Systems.

Information model – A representation of the concepts and relationships for a particular 'domain of discourse'.

Interoperability -

1. The ability to transfer and use information in a uniform and efficient manner across multiple organisations and information technology systems (Australian Government Information Management Office (AGIMO) 2006 p. 3).

2. "The ability for a system or components of a system to provide information portability and inter application, cooperative process control. Interoperability, in the context of the OpenGeospatial Specification, is software components operating reciprocally (working with each other) to overcome tedious batch conversion tasks, import/export obstacles, and distributed resource access barriers imposed by heterogeneous processing environments and heterogeneous data." (Open Geospatial Consortium 2015)

Level of abstraction – the representation of a universe of discourse in a series of models with abstract, higher levels building on more concrete lower level. Levels of abstraction assist in presenting the appropriate level of complexity for particular audiences.

Logical model - a detailed representation of the organisation of data, independent of any particular technology.

Metamodel – a model of a model.

Model Driven Architecture (MDA) - An approach for deriving value from models and architecture in support of the full life cycle of physical, organisational and I.T. systems[59]. The MDA approach represents and supports everything from requirements to business modelling to technology implementations. (Object Management Group 2014)

[59] A "System", in this context, is any arrangement of parts and their interrelationships, working together as a whole. This is inclusive of designs at all levels such as an entire enterprise, a process, information structures or I.T. systems.

Ontology –model comprising the entities, definition of the types, properties, and interrelationships of the entities that exist for a particular domain of discourse (Wikipedia contributors 2015)

Platform Specific Model (PSM) - a model of a software or business system that is linked to a specific technological platform (e.g. a specific programming language, operating system, document file format or database). Platform specific models are indispensable for the actual implementation of a system. (Wikipedia contributors 2015)

Platform Independent Model (PIM) - a model of a software system or business system, that is independent of the specific technological platform used to implement it. (Wikipedia contributors 2015)

Register – A set of files containing identifiers assigned to items with descriptions of the associated items. hierarchical register - a structured set of registers for a domain of register items, composed of a principal register (that contains a description of each of the sub-registers) and a set of sub-registers that contain items from a partition of a domain of information. (ISO 2004)

Register manager - organisation to which management of a register has been delegated by the register owner. (ISO 2004)

Register owner – An organisation that establishes a register. (ISO 2004)

Registration – The assignment of a permanent, unique, and unambiguous identifier to an item (ISO 2004) Registry - information system on which a register is maintained. (ISO 2004)

Registry manager – A person or an organisation responsible for the day-to-day management of a registry. (ISO 2004)

(Geo)Spatial Data Infrastructure (SDI/GDI) –*A* collaborative, approach to creating, maintaining, providing and using distributed geospatial resources under heterogeneous ownership and operation, to meet common information needs. (Box 2013)

(Geo)Spatial Data Infrastructure – the technology, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data (The White House 2002)

Spatial Information Infrastructure – see Spatial Data Infrastructure

Standard - A documented agreement between providers and consumers, established by consensus, that provides rules, guidelines, or characteristics ensuring materials, products, and services are fit for purpose. (OGC, ISO TC211 et al. 2014 p. 5).

Submitting organisation – An organisation authorised by a register owner to propose changes to the content of a register (ISO 2004)

Universe of discourse - see domain of discourse.

Web Feature Service (WFS) - represents a change in the way geographic information is created, modified and exchanged on the Internet. Rather than sharing geographic information at the file level using File Transfer Protocol (FTP), for example, the WFS offers direct fine-grained access to geographic information at the feature and feature property level. WFSs allow clients to only retrieve or modify the data they are seeking, rather than retrieving a file that contains the data they are seeking and possibly much more. That data can then be used for a wide variety of purposes, including purposes other than their producers' intended ones. (Open Geospatial Consortium 2015)

References

ANZLIC—the Spatial Information Council (2014). The Australian and New Zealand Foundation Spatial Data Framework -Making common foundation spatial data ubiquitous across Australia and New Zealand

Canberra, The Commonwealth of Australia (Department of Communications) on behalf of ANZLIC—the Spatial Information Council.

Australian Government Information Management Office (AGIMO) (2006). Australian Government Information Interoperability Framework.

Box, P. (2013). The Governance of Spatial Data Infrastructure: A Registry Based Model Infrastructure Engineering Melbourne, University of Melbourne. Masters of Applied Science (Geomatics) by Research

Box, P. and R. Atkinson (2011). AHGF (Geofabric) production – an analysis of processes, descriptions and models CSIRO: Water for a Healthy Country National Research Flagship. Canberra, Australia, CSIRO: Water for a Healthy Country National Research Flagship.

Brodaric, B. and M. Gahegan (2006). "Representing geoscientific knowledge in cyberinfrastructure: Some challenges, approaches, and implementations." Geological Society of America Special Papers 397: 1-20.

Coetzee, S., et al. (2011). Configuration management of a system of interdependent standards. Standardization and Innovation in Information Technology (SIIT), 2011 7th International Conference on, IEEE.

ed. Prokhorov, A. M. (1983). Domain of discource. Great Soviet Encyclopedia. A. M. Prokhorov. London, Collier Macmillan.

European Commission and European Environment Agency (2014). Midterm evaluation report on INSPIRE implementation. Joint EEA-JRC report. EEA Technical report. Luxembourg, European Environment Agency,

European Union.

ISO (1999). ISO 19103 Geographic information - Part 3: Conceptual schema language, International Organization for Standardization (ISO).

ISO (2002). ISO 19108:2002 Geographic information - Temporal schema, International Organization for Standardization (ISO).

ISO (2003). ISO 19107:2003, Geographic information — Spatial schema, International Organization for Standardization (ISO).

ISO (2003). ISO 19115:2003 Geographic information — Metadata, International Organization for Standardization (ISO).

ISO (2004). ISO 19106:2004 Geographic information - Profiles, International Organization for Standardization (ISO).

ISO (2004). ISO 19135:2005 Geographic information -- Procedures for registration of items of geographic information, International Organization for Standardization (ISO).

ISO (2005). ISO 19109:2005 Geographic information - Rules for application schema, AS/NZS.

ISO (2005). ISO 19110:2005 Geographic information - Methodology for cataloguing, International Organization for Standardization (ISO).

ISO (2005). ISO 19119:2005 Geographic information - Services, International Organization for Standardization (ISO).

ISO (2005). ISO 19123:2005 Geographic information - Schema for coverage geometry and functions, International Organization for Standardization (ISO).

ISO (2005). ISO 19128:2005 Geographic information - Web map server interface, International Organization for Standardization (ISO).

ISO (2007). ISO 19136:2007 Geographic information - Geography Markup Language (GML), International Organization for Standardization (ISO).

ISO (2007). ISO 19136:2007, Geographic information -- Geography Markup Language (GML), International Organization for Standardization (ISO).

ISO (2008). AS/NZS ISO 19131:2008 Geographic information - Data product specifications, AS/NZS.

ISO (2011). ISO 19156:2011 Geographic information: observations and measurements. Geneva, Switzerland, International Organization for Standardization (ISO).

ISO (2014). ISO 19101-1:2014 Geographic information - Reference model - Part1: Fundamentals, International Organization for Standardization (ISO).

Merriam-Webster (2015). Domain. Merriam-Webster Online Dictionary, Merriam-Webster.

Object Management Group (2014). Model Driven Architecture (MDA): MDA Guide rev. 2.0, Object Management Group,.

OGC, et al. (2014). A Guide to the Role of Standards in Geospatial Information Management New York.

Open Geospatial Consortium (2015). "OGC Glossary." Retrieved February 2015, 2015, from http://www.opengeospatial.org/ogc/glossary.

Rumbaugh, J., et al. (2004). The Unified Modeling Language Reference Manual. Boston, Addison-Wesley.

Snell, J. and J. Carpenter (2013). Future trends in geospatial information management:, United National Global Geospatial Information Management Initiative.

The White House, O. o. M. a. B. (2002). Circular No. A-16 Revised. O. o. M. a. Budget. Washington, US, The White House, Office of Management and Budget

Wikipedia contributors (2015). "Abstraction." Retrieved 30-01-2015, 2015, from http://en.wikipedia.org/wiki/Abstraction.

Wikipedia contributors (2015). Ontology. Wikipedia The Free Encyclopedia.

Wikipedia contributors (2015). Platform Independent Model Wikipedia The Free Encyclopedia.

Wikipedia contributors (2015). Platform Specific Model. Wikipedia The Free Encyclopedia.

CONTACT US

t 1300 363 400
+61 3 9545 2176
e enquiries@csiro.au
w www.csiro.au

YOUR CSIRO

Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills for building prosperity, growth, health and sustainability. It serves governments, industries, business and communities across the nation.

FOR FURTHER INFORMATION

Land and Water Flagship
Paul Box
t +61 2 9325 3122
m +61 406 256 006
e paul.j.box@csiro.au
w www.csiro.au