Sub-location N-Dimensional Mixed Category Disaggregation for Energy Efficiency Modelling

K.-K. Yum, G. Foliente, A. Higgins and S. Seo

CSIRO Ecosystem Sciences
Email: Kwok-Keung.Yum@csiro.au

Abstract: In the energy end-use sectors (residential, commercial, industrial), there are significant problems of energy data fragmentation, quality, access and interoperability that make it very difficult to really understand the energy consumption landscape and to develop ways of improving efficiency and/or reducing greenhouse gas (GHG) emissions. This paper presents initial efforts towards the development of an ISO/TC 211 standard based encoding for energy efficiency analysis, making interoperability between key players, so that more effective analysis can be based on wider range of temporal spatial data.

In the standard of Geography Markup Language (GML), the concept *coverage* is used to represent a mapping from a domain (representing a variety of temporal spatial features) to a range of measures (e.g. solar irradiance). Two-dimensional coverages can be conveniently used to represent grid data for geometric features with x- and y-coordinates (e.g. the irradiance mapping over an Australian grid dataset).

There is no GML concept or data structure that have been used in the energy efficiency area.

When modelling energy efficiency, the conventional geometry-based coverage model is not sufficient for the modelling of energy consumption by building type. Here, the new requirement is that energy data are collected or modeled around commercial building types. Buildings are categorised by size (low, medium, high-rise, $l = 3$), climate zone (zone 1 to 7, $m=7$), and building energy performance (star rating 0 to 5, $n=6$). Energy consumption per building EC_{lmn} is simulated and averaged for buildings of size l, in climate zone m and with energy star n. Related sector data are building stock BS_{lmn} of size l, in climate zone m and with energy star n. When encoding these datasets (EC_{lmn} and BS_{lmn}) into a computer model, we created a sub-location mixed category type to divide the disaggregation into an N-dimensional cube (N=3 with l x m x n cells). Each element in the cube (3 x 7 x 6 = 126 of them) represents a distinct building type. These disaggregated items are all inside the value block of the coverage serving as the container of the sub-location disaggregation. As a result, they form a legitimate subdivision of domain features.

The development and use of sub-location N-dimensional mixed category disaggregation type will demonstrate its value and contribution towards the establishment of Energy Efficiency Markup Language (Energy Efficiency ML) schemas along the following aspects: at all levels of geospatial detail, consistent use and understanding of energy consumption data, identifying the most cost-effective energy efficiency and/or GHG emission mitigation measures, and projecting future scenarios. Energy Efficiency ML, based on OGC standards such as Geography Markup Language (GML), Sensor-web observation services (SOS), etc., will be a critical component of a national Energy Efficiency Modelling and Analysis Platform (EE-MAP) (DCCEE 2010). Initially applied to the commercial and residential sectors, this will be eventually extended to all major energy end-use sectors. Similar to the development of the Water Markup Language (WaterML) (WIRADA 2011), which is also based on OGC standards, we will work with national stakeholders and international partners to develop Energy Efficiency ML. Successful Energy Efficiency ML development will allow: (a) problem decomposition into common, fundamental information processing elements, (b) incorporation of advanced data analysis capability, and (c) development of multiple end user applications in the energy end-use sectors aimed towards different audiences. This will enlarge the user base of OGC standards, and deliver high quality data access and compliance.

Keywords: *Energy efficiency, energy markup language, coverage, sub-location disaggregation*

1. INTRODUCTION – KEY ISSUES OF COMMUNICATING ENERGY EFFICIENCY

Energy efficiency is an approach to the processes of consumption and production that reduce energy consumption in the respective process. It offers a great potential to enhance economic prosperity. Increased uptake of cost-effective energy efficiency technologies and processes will help businesses maintain competitiveness towards a low carbon economy. However, due to data fragmentation, mixed use of system boundaries, limited access and lack of interoperability, the information about energy efficiency is not easy to get across accurately to the audience in other domains. In the following we summarise three major issues that we intend to address in energy efficiency modeling.

Issue one: The improvement over energy efficiency is instrumental to addressing many concerns that are related to energy security, climate change adaptation, environmental impacts, productivity and production cost, nationally and internationally. As a result, energy efficiency measures and their computation would be preferred to be available to a variety of other stakeholders so that their concerns could be properly modeled and analysed under their domain contexts. However, this is easier to say than done because the common domain concepts and data of energy efficiency have not been exposed to and shared with other domain experts. This limits the communicative power of the energy efficiency models significantly.

Issue two: There are various measures of energy efficiency for different purposes. Given the number of different jurisdictions, ownerships and industrial processes, the variety of boundary definitions in the energy efficiency calculation can be very large. For example, in one company, the collected waste heat is sold to a third party, and hence is regarded as an output; whereas in another company, the collected waste heat is used within the plant gate, and hence is regarded as an efficiency gain (Tanaka 2008). When the energy efficiency of different nations, regions and companies are compared, their system boundaries and elements within these boundaries can be easily overlooked and the conclusion drawn may be misleading.

Issue three: Within an administration or company, the energy efficiency information may be aggregated from sub-ordinate locations, sites and processes which may also involve different system boundaries due to different cultures and practices. For an efficient assembly of total energy efficiency information, standard could be maintained across the organization so that data can be delivered and assembled correctly. When energy data standards and system boundaries are not explicitly enforced – which is probably true for many administrations and companies – the assembly of energy data is not efficient at the best. Often it can lead to a misleading conclusion.

2. SOLUTIONS

In order to improve the communication capacity of energy efficiency, we propose an Energy Efficiency Measurement and Analysis Platform (EE-MAP) that adopts the following interdependent solutions:

Solution one: Developing ISO/TC211 standards to expose common concepts and data of energy efficiency so that they can be used to provide energy information in the energy or other domains and formulate interactions between the energy efficiency domain and other domains. This encourages the re-use of energy efficiency model elements for external policy modelling and effectively improves the communication between energy efficiency modelling and external domains.

Solution two: Tightening the coupling between energy efficiency data and models with clearly defined terms, concepts and system boundaries so that measures and data are specifically designed for individual purposes of energy efficiency policy development. This enforces the importance of the purpose of the model and its system boundaries rather than energy efficiency numbers generated from the model.

Solution three: Formatting energy efficiency measures and raw data into standard OGC XML schema formats so that they are ready for the execution of the model and the delivery of results. This helps the quick assembly and delivery of the results.

The EE-MAP is intended to contribute to the following aspects of energy efficiency innovation:

- Provide useful data and analysis models for energy efficiency innovation through the integrated consideration of energy efficiency policy, engineering solution and user responses
- Develop common data modelling templates, collectively called the Energy Efficiency Markup Language (Energy Efficiency ML), for the specification of data structures and contents in specific application contexts
- Develop metadata for the easy access to the published products of data and applications
- Facilitate data integration across energy efficiency sectors
- Improve the access to and interoperability of data
- Deliver high quality on data access and compliance.

In the longer term, with the help from national key stakeholders and international collaborators, we are targeting the full-fledge development of EE-MAP to serve the whole government and industry DCCEE 2010).

3. STRUCTURE AND COMPLIANCE OF ENERGY EFFICIENCY ML

Energy Efficiency ML is a set of common energy efficiency model elements that can be used to build a variety of concepts, models, data and applications under the context of Australian and international energy efficiency policies and programmes. Most of these common model elements do not exist at present, but many of them will be derived from existing standards.

3.1. Energy Efficiency ML is XML-based

The common energy efficiency model elements in Energy Efficiency ML will be defined in XML schema format (W3C 2000a, 2000b. and 2000c). These schemas can be then used to encode energy model and data instances under the specific context of applications in the EE-MAP.

3.2. Energy Efficiency ML is based on the ISO/OGC standards

Since common energy efficiency models considered in Energy Efficiency ML are intrinsically related to geographic locations (by name, by longitude and latitude, by code, or by GPS location, etc.), the development of Energy Efficiency ML schemas gains the most benefit by deriving their concepts from the existing Geography Markup Language (GML) standard (ISO 19136) and other dependent OGC standards, e.g. Observation and Measurement (O&M) (ISO DIS 19156 2010), and Sensorweb Observation Service (OGC 2010). The scope of Energy Efficiency ML modeling is focused on developing schemas for the core business concepts of energy efficiency measurement and analysis. Basic temporal spatial data are left to the GML functionality.

3.3. Reusing and extending established feature based concepts and schemas

Energy Efficiency ML will extend the GML feature concept into energy efficiency concepts such as ownership, company, site, generation unit, location, time and time range etc. The GML coverage concept will be used to develop energy related data grid data and time series. The GML/OM observation concept will be extended to cover supply network, end-use demand network, meter and sub-meter network, etc.

4. CONTENT DEVELOPMENT OF ENERGY EFFICIENCY ML

The aspect of Energy Efficiency ML reported in this paper was limited to the key concept of sub-local disaggregation as an illustration of the ISO/OGC approach to data development. The bulk of the work was around the EE-MAP feasibility. A complete development of the Energy Efficiency ML will be considered in the future.

The development of data contents for Energy Efficiency ML originates from the process models and business models from the respective sectors. Figure 1 shows the traditional way of aggregating energy consumption from the energy use pyramid of the IEA (1997). The figure shows two major approaches to energy analysis. The left hand side indicates the top down approach which uses national data of energy balances, accounts and census, etc. The right hand side is bottom up analysis, which may involve new innovative approaches to data surveying and aggregation.

The pyramid shows a hierarchy of energy indicators, from the least detailed to the most detailed. The top indicators may be specified as the energy per GDP dollar, or as energy use per output unit in each sector. The second row may represent a breakdown by sub-sector activities (i.e. typical end uses in each subsector), leading to an indicator of energy intensity measured by energy use per output unit of sub-activities in each sector. These can be further down indicators of more details. The third row represents a further breakdown by equipment types; and the fourth row represents a more refined breakdown into processes within equipment. The figure shows only to the fourth level, which conceptually might not be the final row.

4.1. Sub-location mixed category disaggregation

If we limit our analysis to the sub-sector components, i.e. the second row of the pyramid (Figure 1), the total energy consumption of the residential sector (for a fixed location at a fixed time) can be decomposed into its end use components:

$$E = \sum_{i=1}^{n} A \cdot \frac{A_i}{A} \cdot \frac{E_i}{A_i} = A \cdot \sum_{i=1}^{n} S_i \cdot I_i$$

where i is the index of subsectors (i.e. the index of end use components: i = 1 to 5 for SpaceHeating, SpaceCooling, WaterHeating, Cooking, and Appliance); A is the total activity for the sector – for the residential sector, this is the total population of the location; $S_i = A_i/A$ is i-th structure component, which is the ownership of i-th subsector equipment (space heater, etc.) per capita; $I_i = E_i/A_i$ is the i-th energy intensity component, which is the i-th subsector energy consumption per equipment.
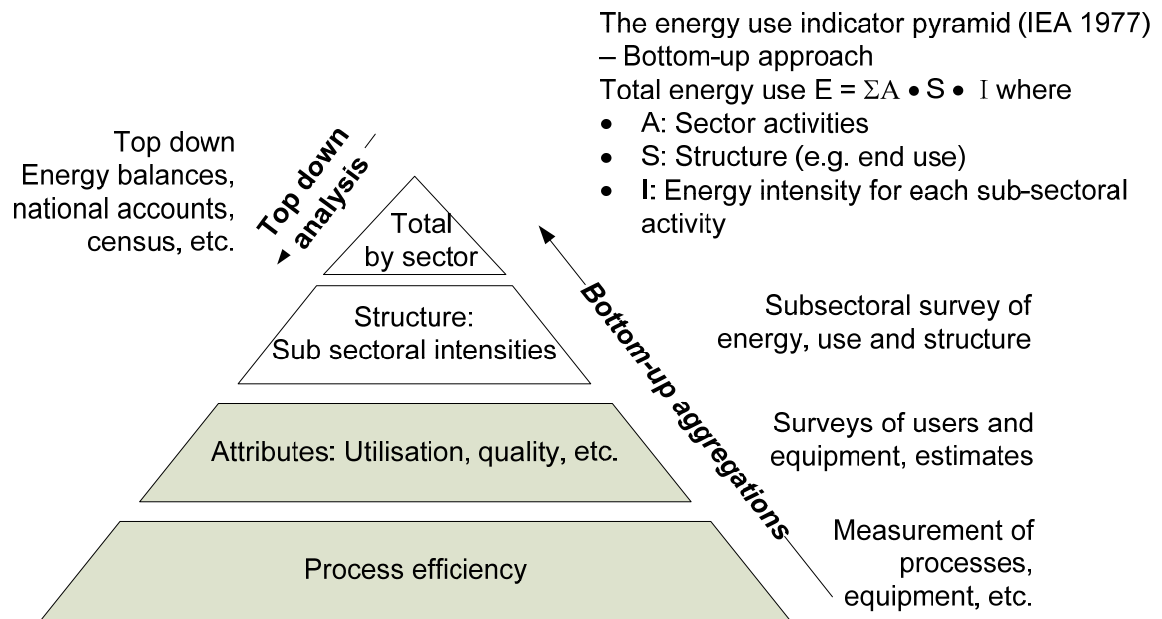

Figure 1. Energy use indicator pyramid where only two (unshaded) levels of details are being used.

Many research projects have been interested in the change in energy consumption as a sum (or product) of the changes in sector activity, structure and intensity – e.g. see Petchey (2010). However, in Energy Efficiency ML modelling, our interest is on the decomposition on the end use components (i.e.: SpaceHeating, SpaceCooling, WaterHeating, Cooking, and Appliance). In fact the breakdown defines the system boundary of four indicators: Total energy, Activity, Structure and Energy Intensity, meaning that any measures not related to SpaceHeating, SpaceCooling, WaterHeating, Cooking, or Appliance will be outside the boundary.

Here the breakdown into sub-sectors is considered within the temporal spatial constraint initially established for these indicators. As such, the breakdown is a sub-location division. It is also a mixed category disaggregation because the breakdown can be on mixed category types of end use.

In ISO/OGC GML, coverages are used to represent variation of measures (e.g. solar irradiance) over a range of features (e.g. a grid of Australian locations). There is no GML concept or data structure that represents sub-location disaggregation like the above end use breakdown of SpaceHeating, SpaceCooling, WaterHeating, Cooking, and Appliance.

To represent a one-dimensional breakdown of end use components within a coverage structure, we can extend the <value>…</value> segment of the coverage schema to include a <MixedCategory> … </MixedCategory> structure for the sub-location mixed category disaggregation. Figure 2 shows a coverage instance including a sub-location disaggregation by residential end-use of energy.

Sub-location disaggregation can be very useful in representing composite index components such as consumer price index components, air quality components, etc. It provides an explicit declaration of the data boundary of the system under consideration.

The standard information modelling method would include a UML model for the information objects required in the application. Here we go straight into the XML encoding (implementation) for the sake of

easy demonstrating. For a complete development in the future, UML diagrams will be supplied, and best practice convention for GML schemas will be followed.
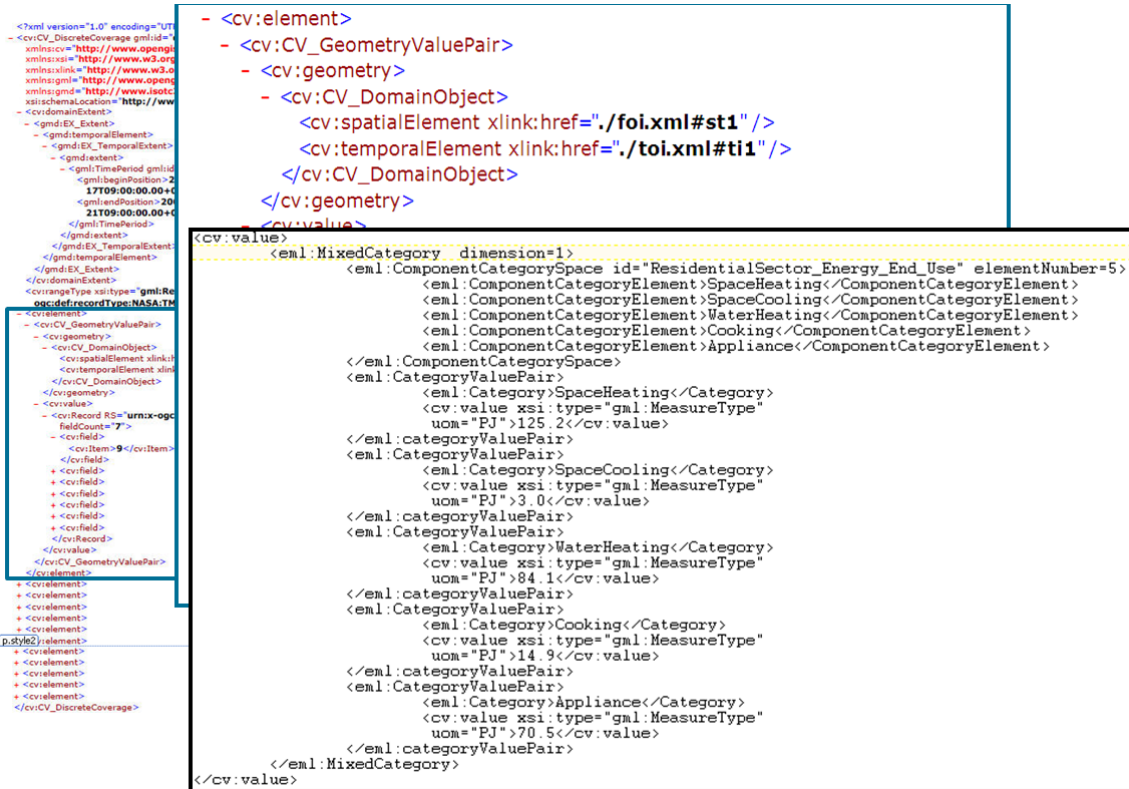

Figure 2. A coverage instance where the value of total energy consumption at space (defined at foi.xml#st1) and time (defined at foi:xml#ti1) is disaggregated by residential energy end use.

4.2. Sub-location N-dimensional mixed category disaggregation

Sub-location disaggregation is not limited to a one-dimensional breakdown as discussed above. It can be disaggregated on N-dimensional ($N \geq 1$) characteristics, such as building stocks characteristics. In a CSIRO Zero Emission Office (ZEO) project, for each local government area (LGA) at any particular year, energy data are collected and modeled around commercial building types. Buildings are categorised by size (low, medium, high-rise, $l = 3$), climate zone (zone 1 to 7, $m=7$), and building energy performance (star rating 0 to 5, $n=6$). Energy consumption per building EC_{lmn} is simulated and averaged for buildings of size l, in climate zone m and with energy star n. Related sector data are building stock BS_{lmn} of size l, in climate zone m and with energy star n. So the commercial building sector energy consumption at any particular LGA and year t is:

$$E_{COM,} = \sum_{l=1}^{3} \sum_{n=1}^{7} \sum_{n=0}^{5} EC_{lmn} . BS_{lmn},$$ where all values are varied according to the temporospatial location at the LGA in year t.

When encoding this equation into a computer model, we created a sub-location mixed category type. The instance data are similar to the one in Figure 2, to divide the sub-location breakdown into an N-dimensional cube (N=3 with l x m x n cells). Each element in the cube (3 x 7 x 6 = 126 of them) represents a distinct building type.

Local government areas are administered by different local governments. If they all agree to collect the same set building stock information BS_{lmn} (with respect to size l, climate zone m and energy star n), then their commercial building sector energy consumption E_{COM} can be easily simulated and effectively assembled at various temporospatial location scales. The availability of building stock data, in the format of sub-location N-dimensional mixed category, helped us simulate and aggregate the commercial building energy consumption for all LGAs in NSW effectively (Figure 3).

Finally, it must be pointed out that there are commercial and privacy concerns about releasing detail building information for all individual buildings; information about building stock type is probably a level of approximation that both the industry and government may be comfortable with .
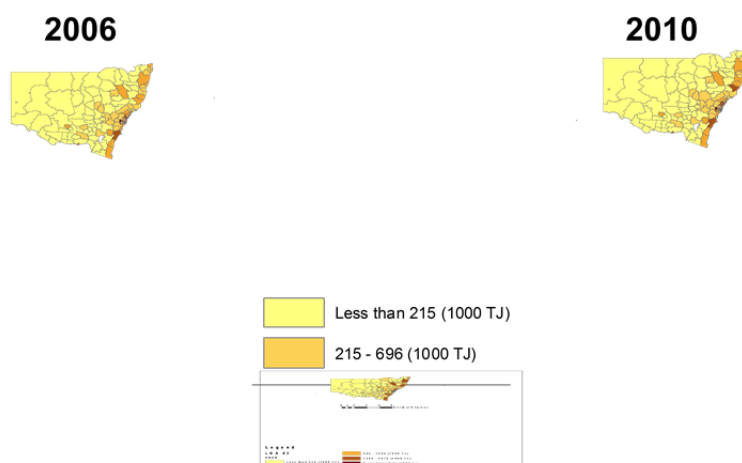
Figure 3. NSW energy consumption of commercial buildings by LGA in year 2006 (based on existing building stock) and year 2010 (based on simulated building stock)

5. WHAT WE ARE AND WHAT WE ARE NOT

Our goal is to develop ISO/TC 211 standard based energy efficiency schemas. There is no other initiative doing that. Our overall approach is similar to the approach of Water Information Research And Development Alliance (WIRADA 2011), except that we are in a different domain.

The following is what we are not.

Building information model (BIM, http://www.buildingsmart.com) and CityGML (http://www.citygml.org/) target the modelling of *individual* buildings. They are closely related to the Intelligent Use of Buildings' Energy Information (INtUBE, http://www.intube.eu/) and related projects (http://www.bimtoolset.org/ontologies/). Our work is focused on making available the complete energy efficiency information about the whole geospatial area at various granularity levels (but not to the level of individual building). Although BIM applications can be used for modelling detail energy consumption for individual buildings, there is no way we can get information of *all* individual buildings together, due to commercial and privacy concerns.

Semantic interoperability of geospatial datacubes (Sboui et al 2007) focuses on the interactive analysis of multidimensional datacubes, i.e. a kind of Spatial On-Line Analytic Processing (SOLAP), often for data-mining purposes. Our work is around the development of transaction systems for the energy efficiency measurement and analysis at various levels of geospatial detail.

Statistical Meta Data eXchange (SMDX, http://www.sdmx.org) identifies a standard technique for modelling, expressing, and understanding the structure of this multi-dimensional "cube", allowing automated processing of data from a variety of sources. SDMX Information Model provides for the structuring not only of statistical data, but also of "reference" metadata; whereas, Our EE-MAP methodology provides for the structuring of energy-efficiency applications, together with data standardised and formatted for transactional efficiency.

Despite all these differences, there are a lot to learn from these alternative frameworks and methods.

6. BENEFITS AND COSTS OF EE-MAP

As a part of the feasibility analysis of the EE-MAP, we can draw up the following list of benefits of our EE-MAP approach:

- Reusing and refining established feature based concepts and schemas help enhance precise communications and reduce reinventing the wheel.
- Standard abiding enables effective assembling and delivery of observation and simulation data/results to collaborative partners (through Standard services such as sensorweb observation service).

- Encouraging effective partnerships with industry, research and local community in the co-development of data, tools and research results.

But these benefits are obtained at a cost that can be listed as follows:

- Exposing common concepts and data goes against the grain of scientists who wants to keep data, tools and (raw) research results to themselves; at present there is not enough motivation for the scientists to share data.
- The harmonisation of data and protocol standards takes time and effort.
- The complexity of various GML schemas can be daunting to non-GML specialists.
- There are not enough catalogue tools, learning tools and transformation tools to help end users to read, understand, use and even develop schemas of their interest.

7. WAYS TO MOVE ON

Like the development of science and engineering in the large, the development of the EE-MAP cannot and should not be confined within the boundary of any individual research institution. We propose to develop Energy Efficiency ML, together with other initiatives such as WaterML, as a part of the whole of nation collaborative research infrastructure initiative. This will create much more end-users than individual desktop applications can generate. As such, new commercialisation and legal frameworks for sharing data and tools will help extend the national and international research infrastructure.

We may collaborate with organizations such as ISO TC211/OGC to delegate the harmonization effort to them.

We may want to develop a national body of energy efficiency research and information alliance to

- attract researchers to participate in Energy and EE-MAP development, creating more examples and real data for use;
- develop catalogue tools, learning tools and data transformation tools to help the participation from domain experts.

ACKNOWLEDGMENTS

REFERENCES

DCCEE (2010), PM Task Group on Energy Efficiency Report, DCCEE, Canberra, ACT.

IEA (1997) Indicators of Energy Use and Efficiency: Understanding the link between energy and human activity. Paris: International Energy Agency.

ISO 19123 (2005) Geographic Information — Coverages

ISO 19136 (2007) Geographic Information — Geography Markup Language.

ISO DIS 19156 (2010) Geographic Information — Observations and Measurements (OGC Abstract Specification Topic 20). OGC Reference number: OGC 10-004r3.

OGC (2007) Observations and Measurements – Part 1 – Observation schema. OGC Reference number 07-022r1.

OGC (2010) OGC web services common standard. OGC Reference number 06-12r19

Petchey, R. (2010) End use energy intensity in the Australian economy, ABARE-BRS research report 10.08. Australian Bureau of Agricultural and Research Economics, Bureau of Rural sciences.

Sboui, T., Bedard, Y., Brodeur, J. and Balard, T. (2007) A conceptual framework to support semantic interoperability of geospatial datacubes. ER workshops 2007, (eds.) Hainaut, J.-L., LNCS 4802, pp. 378-387.

Tanaka, K. (2008) Assessing measures of energy efficiency performance and their application in industry, IEA Information Paper, International Energy Agency, Paris.

W3C (2000a) "XML Schema Part 0: Primer", W3C, April 2000. (http://www.w3.org/TR/xmlschema0)

W3C (2000b) "XML Schema Part 1: Structures", W3C, April 2000. (http://www.w3.org/TR/xmlschema1)

W3C (2000c) "XML Schema Part 2: Datatypes", W3C, April 2000. (http://www.w3.org/TR/xmlschema2)

WIRADA (2011) Water Information Research And Development Alliance. A joint initiative of the Bureau of Meteorology and CSIRO. Accessed in August 2011 http://www.bom.gov.au/water/regulations/wdtf/wirada.shtml