PST201309231953371GE

The Big Data Format Service on Energy Usage Profile with ISO Standard

Chung-Ming Own
School of Electronic and Computer Engineering
Georgia Institute of Technology
Email:chung.ming.own@me.com

Ming-Whei Feng
Smart Network System Institute (SNSI)
Institute for Information Industry
Email:mfeng@iii.org.tw

Da-Sheng Lee
Department of Energy and Refrigerating Air-Conditioning Engineering
National Taipei University of Technology
Email:f11167@ntut.edu.tw

Emery Jou
Smart Network System Institute(SNSI)
Institute for Information Industry
Email:emeryjou@iii.org.tw

***Abstract*—Big data which has been given rise by the incomprehensibly large worlds of information, is the explosion of mobile networks, cloud computing and new technologies. Besides, big data is a collection of complex data sets, which becomes difficult to process with on-hand database management tools or traditional data processing applications. Thus, the computing of big data has become ubiquitous and impressive. Accompany with the advanced researches on the Internet of Things, more and more user profiles are collected from the Internet, the big data solutions are conspicuous. In our study, the authors focused on the energy conservation and carbon reduction, it is important to devise a realistic method for an overview on the general energy usage and specific electricity needs. This kind of data is not just a back-office, accounts-settling tool any more. This study proposes the energy sensor profiles based on the ISO 19115 and ISO 19119 to realize the metadata standard, which can be used to communicate and share all energy usage data on the production process. According to the results, the industries will benefit from the earnest practice and analysis of the shared energy usage information.**

***Index Terms*—Energy Usage Profiles, ISO 19115, Data format.**

I. Introduction

The volume of data we deal with has grown to terabytes and petabytes in this decade. When the volume of data keeps growing, the types of data generated by the applications are become more complex than before. Hence, traditional technologies are not handy to analyze, search, store and manage these kinds of data. That is the big data. On the other hand, companies across nearly every industry find they not only need to manage increasingly large data volumes in their real-time systems, but also to analyze that information so they can make the right decisions to compete effectively in the market. Hence, a flexible and scalable solutions are needed to handle their big data easily and effectively in such a kind of environment.

Big data requires exceptional technologies to efficiently process large quantities of these data in the tolerable time. In [1] , the author suggests suitable technologies to handle big data include associate rule learning, classification, cluster analysis, crowdsourcing, and data fusion and so on. Furthermore, multidimensional big data represented as tensors can be handled efficiently by the tensor-based computation [2]. According to IDC, it is imperative that organizations and IT leaders should focus on the ever-increasing volume, variety and velocity of information that forms big data [3]. The volume, variety and velocity are the three important aspects to handle big data. Especially in the variety, data comes in all types of format; from traditional databases to hierarchical data stores created by end users and OLAP systems, to text documents, email, meter-collected data, video, audio, stock ticker and any other financial transactions. Sometimes, these organization's data is not numeric about 80%. Thus, a more dynamic, flexible database schema format is needed to handle these structured, semi-structured, and unstructured data that comprises today's big data.

Taiwan is deficient in conventional energy resources and highly dependent on energy imports (nearly 99%), with nearly 90% of it greenhouse gas emissions coming from carbon dioxide emitted from energy use. As a member of the global village, Taiwan has committed itself to sharing the obligation and responsibility. Thus, energy conservation has been a major components in the comprehensive climate protection programs in Taiwan, because the release of carbon dioxide emissions by the utility companies powering the company's facilities and from the use of fuel for heating or cooling represents the greatest potential climate impact. Because of that, a principal focus on energy usage has been the Taiwan's conservation goal.

This study examines methods by which to use energy in an effective and stable manner, as well as strategies for converting energy to assist in companies' growth. For example, fossil-fuel energy is currently the primary source of energy and is converted to other types during the combustion process. This process produces substantial greenhouse gases composed primarily of carbon dioxide and suspended substances involved in air pollution. However, the effectiveness of using fossil fuels is not high. Thus, energy-related questions have recently become emphases in environmental protection and technology issues. The objective of this study is to ameliorate greenhouse effects and resolve the potential energy crises that might be caused by future energy shortages through research and collaboration between various fields.

Because energy monitoring through sensors in Taiwan is still conducted by various government agencies, industrial parks, and private industries, their monitoring formats are incompatible, and data quality control and accuracy are inconsistent, leading to additional manpower and cost consumption to ensure data format conversion and content accuracy. This situation limits the data flow and application sharing, and even results in repeated investments in monitoring resources. Hence, in this study, the authors followed the relevant in ISO 19115 metadata standard and used them to design a shared energy resource format, and coordinated with records to share energy resource metadata items based upon the environment in Taiwan. Demonstrate results not only reveals the need of the big data format, but also explosive the idea of big data to the industry.

Following the introduction, the remainder of this study is organized as follows. Section II introduces the outlook of common data format and ISO 19115. Section III presents a description of the energy usage sharing framework for our study. Section IV provides the description of an expansion of renewable energy in our system. Section V offers a conclusion.

II. Preliminaries

A. ISO Standard 19115 and 19115-2

The primary objective of metadata is to allow users to understand and appropriately use data. In addition, metadata can be the description result of currently stored data or services. According to the planning of the Federal Geographic Data Committee of the United States, metadata can undertake the following tasks: data content discovery, evaluation of data suitability, data access, data use, data transfer and conversion, and data management. In other words, through metadata, people who require data can successfully complete the search of existing data that conform to search conditions, interpret data characteristics that conform to the conditions, evaluate whether the data contents match requirements, obtain information regarding data-acquisition channels, and understand how to convert and use the data.

ISO 19115 is a kind of metadata. The objective of ISO 19115 is to provide general-purpose metadata for describing digital geographic data. This International Standard defines metadata elements, provides a schema, and establishes a common set of metadata terminology, definitions, and extension procedures. More detailed metadata for geographical data types and geographical services are defined in other ISO 191xx suites of standards and user extensions. ISO 19115 contains two metadata data types: extent data and citation and responsible party information. Extent data are an aggregate of the metadata elements that describe the spatial and temporal extents of the referring entity. They contain data about the geographic, temporal, and vertical extents of the referring entity. Citation and responsible party data provide a standardized method for citing a resource and data on the party responsible for a resource.

The objective of the ISO 19115-2 draft is to provide an additional structure to more extensively describe the derivation of geography data, but it is currently not a final norm and it lacks important parameters. An analysis of ISO 19115-2 shows that some entities have been modified in the draft, and the existing geographic metadata standard has been extended by defining schema required for describing imagery and gridded data.

In our model, the metadata entity set is the root package of the metadata, and an additional package of metadata is specified. MI_Metadata is a specified subclass of MD_Metadata. It is used to provide data describing energy data. The extension provided through MI_Metadata adds an association to the MI_AcquisitionData class. MD_Identification is used to take the responsibility on the identifying energy profile provider. MD_AggregateInformation, MD_DataIdentification, and SV_ServiceIdentification are used to provide data on the data aggregation, specific data description, and web service setting.

B. Web Service

According to the definition of the W3C (World Wide Web consortium), a web service is "a software application identified by a URI, whose interfaces and bindings are capable of being defined, described, and discovered as XML artifacts [4]. A web service supports direct interactions with other software agents using XML based messages exchanged via internet-based protocols". This definition implies that communication is done in XML, which is the standard for information exchange through the net.

The essence of Web Services is thus to break the complex task down into simpler ones. By standardizing the interface to software modules, we are able to choose services by solving the same task and reuse existing services for the complex problem solving.

The Service Oriented Architecture (SOA) pattern provides the fundamental roles of Web Service provider and consumer within a distributed computing network [5]. This pattern emphasizes that computing tasks or solutions may be devised from performance of an ad hoc configuration of individual services, about which only the types and holding need to be known. SOA also focuses the basis quests, service responses, or service exceptions.

The fundamental concept consists of three functions: Publish, Find and Bind. A new service offered by a service provider is published to a registry. As part of the registration a service description is provided, in order for other services or client to be able to identify them. This registry is accessed by distributed clients to find services providing specific functionality or information. After identifying a service, the clients bind to it and interact directly. Thus, the registry is just a mediator providing the means to publish and find the services. The system diagram of the web service is shown in figure1.

III. Big Data Format Service

In our system, the implementing rules defined the service interfaces at an abstract level. Thus, at the implementing rule level, the service interface specifications do not specify the implementation technology or standards to be used. Our system imposes only obligations on data sets and services.

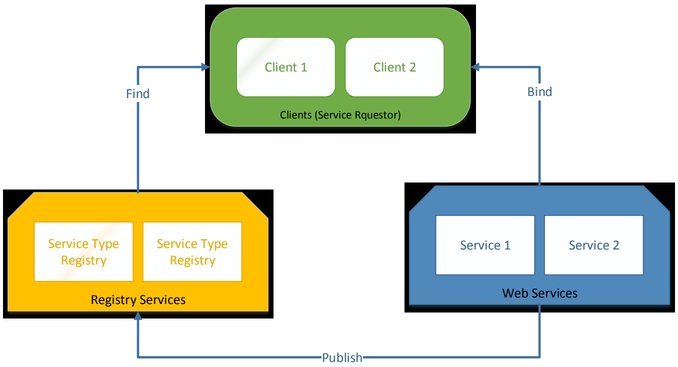

Figure. 1 The system diagram of web service

Discovery Service

The Discovery Services in our system allow users and computer programs to search for spatial data sets and services based on their metadata records. Two types of metadata are to be created:

- Metadata about the data set being published.
- Metadata for the network services being published.

These two types are closely related, and both must be managed together.

Creating and publish metadata

Metadata is created when data or services are created, and thus, the responsibility for creating metadata belongs to the Data Provider. For metadata creation to be deferred has been a frequent tendency, coming later in the lifecycle as part of the production of a directory of data resources. Moreover, publishing data and service metadata into a spatial data infrastructure is the task of the data publisher by using a central registration service. The data publisher can be the data provider or the publisher of the industry that wants to share its energy profile. Publishing is a simple process of registering the URL for the data and service metadata resources into a registry or catalogue service.

View and event report services

The View Service allows users and computer programs to view spatial data sets. The view service provides the ability to display, navigate, zoom, and display historical legends and relevant metadata. Moreover, the Event Report Services report the specific energy data to the registered user regularly. The implementing rules for these services define the operations that the service needs to support and the quality of service criteria that need to be met.

Download service

The Download Service gives users access to the full data content captured and maintained as spatial objects and collected into spatial data sets. The Download Services provide the capability to integrate energy data with other types of business data and to support decision making through a more detailed analysis on the data.

The big data format service

In this study, we divided the big data format service provided by domestic vendors into the following four categories based on the energy service type provided by the vendors:

- Energy monitoring service (EMS)
- Electrical energy supply service (EES)
- Steam supply service (SSS)
- Petroleum and petrochemical service (PPS)

Thus, for these four services, we proposed energy sensor metadata standard specifications similar to the EIP, and used these data specifications to establish an open resource-sharing platform. Under the control of this platform, each case vendor can openly exchange its energy use information, and convert the data into an energy information service that can be read and analyzed easily. Because of confidentiality, the more sensitive contents are hidden.

To describe a network service, a service name, service location, a function provided by the service, a parameter name and type for the input/output required by the function, and other basic descriptions of service characteristics must exist. In addition to these required characteristics, the sensor model, energy type, and time characteristics unique to the energy service must be considered. Therefore, this study followed the definition in ISO 19119, and primarily defined network service in energy information search as the "sv_ServiceIdentification" class for describing the service type, service versions, data obtained by the service, and service limitations. We still need to include the "sv_OperationMetadata" class to primarily records the service connection and operating methods. The fields have a DCP item that records data related to an Internet connection to the server, and an operationName item that records the operating method provided by the service. The service needs to include the "sv_Parameter" and "sv_OperationChainMetadta" classes to describe related parameters and operational chains.

Accordingly, for the purpose to define the big data format service, in our study, we divide the shareable energy information by the following models,

The energy and industry model

In this model, the energy information shared by vendors is related to the industry. In this category, the vendors can provide energy consumption data of diverse equipment in various case plant locations based upon the industry type and the Unified Numbers of the companies. To accurately and automatically search through the energy information of the equipment, we added energy data sensors to the equipment for the plant and sent the data out periodically through the use of wireless short-domain communications protocol technology.

The energy and product model

In this model, the energy information shared by the vendors is related to individual products. The vendor can provide data of energy consumption by the related production equipment based upon the product's EPC code or product name. Similarly, the sensor data inside the equipment must be accurately described.

The energy and produce chain model

In this model, the energy data provided by the vendors are related to the industry production chain. The data provided by the vendor are categorized according to different production chains. Various types of equipment in this category must accurately describe sensor data inside the equipment.

The energy and location model

In this model, the energy data shared by the vendors are related with the location on the plant. The data provided by the vendor are categorized based on different geographical locations. We also expected that this category of data can show the relationship among energy consumption, geographical location, and the plant orientation. In this data category, various types of equipment must also accurately describe the sensor data on the equipment.

Accordingly, to define the format schema that enables participating units (including energy companies, software suppliers, data and service providers, and government agencies) to search and exchange energy data, the authors establishe energy use information, source, time, space and other characteristic data to facilitate the effective search, acquisition, and application of these data, and to enable them to support the formats of different systems. The authors primarily used the MD_Metadata and MD_Identification classes in ISO 19115 and ISO 19119 and their subclasses. The MI_AcquisitionInformation class is used primarily to collect all sensor data, MD_DataIdentification is used to record product information, SV_ServiceIdentification is used to provide different services of the case plants, and MD_AggregateData is used to collect all the detailed data from the sensors. The subclasses of MI_AcquisitionInformation include MI_Platform, MI_Operation, and MI_Instrument. MI_Platform is a service platform used to collect sensor data, and its attributes include citation, identifier, description, sponsor, and instrument, which can be used to record sensor use, brand recognition, remark description, and brand and sensor name, respectively. MI_Operation can be used to record sensor operation data, and its attributes include description, citation, identifier, status, type, parentOoperation, and platform, which can be used to record sensor operation description, operating records, the sensor numbers in the same equipment, sensor calibration status, sensor model number, associated data from the sensors, and the service record of this sensor, respectively. MI_Instrument is used to describe the data items of the sensor equipment that collects the data, including citation, identifier, type, description, and mountedOn attributes used for recording the measurement unit of the sensor, data range, sensor model number, sensor description, and purpose of the sensor measurements, respectively.

The following is an example of the big data format schema. If we want to tag the vendor-provided energy service data with data identification classes, we have to be able to express spatial data content, and also define the data name, data class, and the reference information. Thus, we need to include the following attributes: platform, project, operation, instrument, place, theme, and data center. The example is listed as follows,

```
<gmd:MD_Keywords>
  <gmd:keyword>
    <gco:CharacterString> digitalpowermeter
    </gco:CharacterString>
  </gmd:keyword>
  <gmd:type> instrument </gmd:type>
  <gmd:keyword>
    <gco:CharacterString>Energy Monitoring
Service
    </gco:CharacterString>
  </gmd:keyword>
  <gmd:type>instrument</gmd:type>
  <gmd:keyword>
    <gco:CharacterString>Datasets Collection
Platform
    </gco:CharacterString>
  </gmd:keyword>
  <gmd:type>platform</gmd>
 </gmd:MD_Keywords>
```

IV. The case study of Big data format service

The following is a case study of the big data format service. The Technology Energy Research Center of National Taipei University recently established a corporate energy usage monitoring platform [6]. The objective is to collect and monitor industry energy data. The historic energy consumption data of plants can be obtained using the monitoring platform. After the data are obtained, energy-saving analysis and methods can be provided to the industry as reference data for energy savings, thus achieving energy-saving improvement measures. The vendor that is currently providing energy use data is the Taiwan SKF company [7].

The entire implementation process involves installing digital power meters on the equipment in the Taiwan SKF Company’s plant to understand the electrical consumption situation of the main axle production process and other equipment and machines. The data were returned to the National Taipei University service center for every three minutes.

Figure 2 shows the data interpretation screen after the SKF vendor data were sent to the server. The screen primarily shows the electrical usage calculation used by the equipment during that month and the energy use status of that day. Figure 3 shows the electricity usage by the SKF warehouse equipment, which describes the total power usage of the warehouse; the wattage used by Running-in Station 1, Running-in Station 2, the air compressor, and lathe processing room; the power factor; communication status; and the time of the last data entry.

V. Conclusion

For decades, the economy in Taiwan has been continued to grow, forming a positive linear relationship between the rate of an increase for energy consumption and the economy. Consequently, this study examines methods by which to use energy in an effective and stable manner, as well as strategies for converting energy to assist in positive economic growth. In this study, we design the big data format service to formulate the energy sensor metadata and create a sharing and reuse value of domestic EMS, ESS, FSS, and SSS field energy management data, provide cross-industry maintenance of energy service systems, and to expand domestic energy service-related markets. Besides, the authors propose the energy sensor profiles based on the ISO 19115 and ISO 19119 to realize the metadata standard, which can be used to communicate and share all energy usage data on the production process. According to these results, the industries will benefit from the earnest practice and analysis of the shared energy usage information.

Figure. 2 The energy usage monitoring platform

Figure. 3 The SKF warehouse real-time power usage data

Reference

[1] M. James, C. Michael, B. Jaques, B. Brad, D. Richard, R. Charles, B. Angela Hung, “Big data: the next frontier for innovation, competition, and productivity,” McKinsey Global Institute, May 2011.

[2] “Future directions in tensor-based computation and modeling,” workshop report, May 2009.

[3] C. Philip, “Big data analytics: future architectures, skills and roadmaps for the CIO,” IDC, September 2011.

[4] “Web Services Architecture Requirements,” W3C working draft, April 2002.

[5] “Web Services Architecture,” W3C Working Group Note, February 2004.

[6] The energy usage monitoring platform, http://210.71.198.79:88/SKF.php

[7] Taiwan SKF company, http://www.skf.com/portal/skf_tw/home