

Towards Ontology Based Data Extraction for Organizational Goals Metrics Indicator

Tengku Adil Tengku Izhar[1(✉)] and Bernady O. Apduhan[2(✉)]

[1] Faculty of Information Management, Universiti Teknologi MARA, UiTM Selangor, Shah Alam, Malaysia
tengkuadil@yahoo.co.uk

[2] Department of Information Science, Faculty of Science and Engineering, Kyushu Sangyo University, Fukuoka, Japan
bob@is.kyusan-u.ac.jp

Abstract. In this paper, we proposed a measurement framework to evaluate the quality of results at organizational goals level. We proposed metrics indicators to guarantee and optimize the usage of data to extract useful information in organization. The framework is flexible to change without affecting things around because the framework is applicable in any domains. We discuss the review of the problems, proposed solution to the problems and draw a conclusion of this paper.

Keywords: Big data · Data extraction · Data analysis · Metrics · Ontology

1 Introduction

Organizational goals ontology focus on the usage of organizational data instead of knowledge, information or tools because organizational data is a major resource in every organization and it is important to evaluate the relevance of this organizational data in achieving the organizational goals. We also suggest organizational data is important as information and knowledge to assist the decision-making process [11–13]. In an organization, it is extremely important for the manager to have access to the most relevant organizational data in relation to the organizational goals. It's pointed out that sharing important data and information can provide the required knowledge to assist successful decision-making. It is crucial for organizations to create and generate new data and evaluate it to enhance decision-making. Different ways of generating new ideas, information and knowledge will help in terms of decision-making and will enable teams within the organization to use the most relevant organizational data to successfully achieve the organizational goals [1–5].

2 Literature Review

Most of the recent studies focus on the development of systems ontologies and enterprise ontologies [12, 13]. However, not many studies have been conducted on organizational goals developed an organizational ontology but this study discussed the

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12252, pp. 261–276, 2020.
https://doi.org/10.1007/978-3-030-58811-3_19

development of knowledge mapping based on ontology. Previous study did not evaluate any organizational resources such as data, information or knowledge in their model, rather, they developed an ontology to show the relationship between the organizational role and organizational activity. ontology. There are a very limited number of studies which evaluate organizational resources, such as data, in relation to the organizational goals. An ontology shows the relationship between the knowledge domains within the organization. This domain knowledge is important for any future domain expert and entrepreneur to identify the goal elements in their organization. It is important for them to recognize the relevant data in relation to the organizational goals. In order to evaluate the organizational data, metrics is used. We suggest that a metrics model is important in measuring the extent to which the organization data are consistent with the organizational goals. The purpose of the organizational goals ontology is to improve the understanding of the organizational structure and the relationship between the organizational goals. Recent studies focus on the framework and the integration of ontology [3–5]. These two aspects are very important in the development of every ontology. However, an ontology also needs to focus on relationships and structure, especially in relation to organizational goals. Thus, other aspects that need to be considered are relationships and structure. Relationships and structure are important aspects in the achievement of the organizational goals. Therefore, an ontology is important to improve the common understanding of the structure and relationships in the achievement of the organizational goals.

3 Ontology

In organization, it is extremely important for the manager to have access to the most relevant organizational data in relation to the organizational goals. Sharing important data and information can provide the required knowledge to assist successful decision-making [11]. It is crucial for organizations to create and generate new data and evaluate it to enhance decision-making. Different ways of generating new ideas, information and knowledge will help in terms of decision-making and will enable teams within the organization to use the most relevant organizational data to successfully achieve the organizational goals. Data is presented in many forms such as documents and statistics. These data are the most important resources in relation to the organizational goals.

In this paper, the first step to identify the relevant organizational data is to recognize the matching set of organizational data to identify which organizational data relate to the organizational goals. A large body of existing work uses the terms record linkage, data linkage, record matching, and data matching. We use the term data dependency in an effort to identify the dependency relationship between the organizational data and organizational goals because we attempted to identify dependency for all organizational data that relate to organizational goals (Fig. 1).

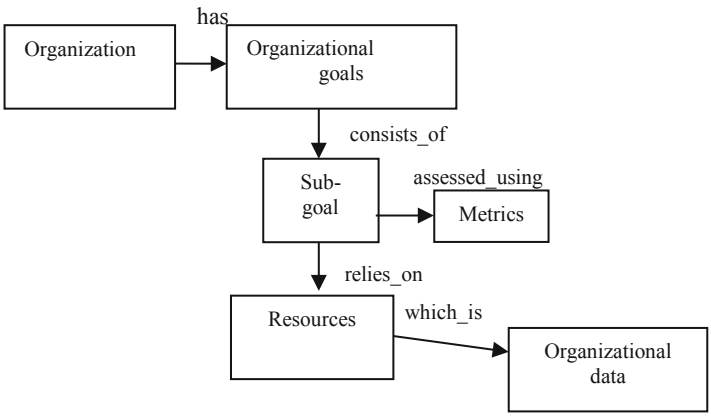

Fig. 1. Organizational goals ontology

3.1 Metric Rules

Metrics models are designed to address organization process which include organization objective and to assist the decision making process. Metric must be clarified.

Metrics. A verifiable measurement used to measure both quantitative and qualitative. As the volume of data increase, metric provide data refinement.

Metrics Requirement. Metric design of what need to be accomplished during the metrics process. Specific team attempts to identify the needs toward organization objective.

Verifiable: Set of data that has been agreed for converting process into measure.
Measure: Characteristics in a numerical or nominal form.

Metrics Analysis. Requirement must be fulfilled.

Control: Metrics enable to evaluate and control the source they are measures.
Communication: Communicate externally and internally for the purpose of control.
Improvement: Identify gaps for improvement.

Let's denote that metrics requirement clarify as verifiable and measure as $V_{erifiable}$ and M_{easure}

$$MetReq : (V_{erifiable}, M_{easure}) \tag{1}$$

and denote control, communication and improvement as C_{ontrol}, $C_{ommunication}$ and $I_{mprovement}$.

$$MetAna : (C_{ontrol}, C_{ommunication}, I_{mprovement}) \tag{2}$$

Above may also be written as

$$Met : (MetReq)(MetAna) \tag{3}$$

so,

$$Met : \left(V_{erifiable}, M_{easure}\right)\left(C_{ontrol}, C_{ommunication}, I_{mprovement}\right) \tag{4}$$

where *MetReq* and *MetAna* are the variables of *Met* which $V_{erifiable}$ and M_{easure} are the characteristics for *MetReq* and C_{ontrol}, $C_{ommunication}$ and $I_{mprovement}$ are the characteristics for *MetAna*.

3.2 Data Concepts

Collection of raw resources and converting this data into useful information. Organization relies on data in every aspect of their actions. It is important for organization to identify, create, store and analyse this data. Data must be clarified.

Data. Organization relies on this data toward organization objective and action. Specific team use data and convert it into valuable information.

Quality Data. Data must be in parallel with the organization's needs. Data in organization need to be subject-oriented such component of subject matter and improve the effectiveness in responding to queries. Thus, data must be,

Complete: Complete presence of the corresponding data records. Data completeness refer to the lack of needed fields.
Accurate: Data is correct and set a context for further analysis.
Current: Data is up to date to organization need.

We denote that complete, accurate and current as $C_{omplete}$, $A_{ccurate}$ and C_{urrent}. Thus,

$$Data : (QD) \tag{5}$$

so,

$$Data : \left(C_{omplete}, A_{ccurate}, C_{urrent}\right) \tag{6}$$

where data must be quality to support future need and performance. Data must be $C_{omplete}$, $A_{ccurate}$ and C_{urrent} in order to become quality.

3.3 Organizational Process

Organization process ($\text{Org}_{\text{process}}$) can be partitioned into several processes. However, organization process can be very large, so in this paper, the focus of organization process is on their existing data. The algorithms are defined based on the organization goals and to show the entire relationship of the organization goals rather than looking at the relationship between business side and data side. Our approach is to look at the

organization data flow and the impact of data toward organization goals. Our approach is also intended to look at the relationship between the organization goals.

Process Input. During processing, the involved data called process input (P_i), where $\rightarrow$ is an involvement process. So decide,

$$\mathrm{Org}_{\mathrm{process}} \rightarrow \mathrm{P_i} \tag{7}$$

Process Output. Every process generate output, so decide which organization process involve process output (P_o) where $\rightarrow$ also is an involvement process.

$$\mathrm{Org}_{\mathrm{process}} \rightarrow \mathrm{P_o} \tag{8}$$

3.4 Data Process

Data process ($D_{process}$) can also be partitioned into several process. Here, process shows the flow of data within the organization.

Data Input. Every stage of the process involves data input. Every organization created data almost every day and this data is stored in data storage such as database. So data process involve data input (IN).

$$\mathrm{D}_{\mathrm{process}} \rightarrow \mathrm{IN} \tag{9}$$

Data Output. Process generated output. Data stored in data storage need to be evaluated to make sure that the data is valuable for certain action. So decide which data process involve data output (OT)

$$\mathrm{D}_{\mathrm{process}} \rightarrow \mathrm{OT} \tag{10}$$

Let, $\mathrm{Org}_{\mathrm{process}} = \mathrm{P_i}, \mathrm{P_o}$, where ϵ is a characteristic x can find in P_i. For example y is an output of x

$$\mathrm{Org}_{\mathrm{process}} = \left\{ \mathrm{Org}_{\mathrm{process}}(x, y) | x \in \mathrm{P}_i, y \in \mathrm{P_o} \right\}. \tag{11}$$

Then organization process can summarize as

$$\mathrm{Org}_{\mathrm{process}} = \left\{ \mathrm{Org}_{\mathrm{process}}(\mathrm{IN}, \ \mathrm{OT}) | \ \mathrm{IN} \in \ \mathrm{P_i}, \mathrm{OT} \in \mathrm{P_o} \right\}. \tag{12}$$

It shows the entire $Org_{process}$ rely on *IN* and *OT*. Here, P_i rely on *IN* and P_o rely on *OT*. The process define as y depend on x and it concludes as *OT* rely on *IN*. Objective and action are the main requirement to achieve the develop target within the organization. Develop target is organization main goals. Objective here can be assumed as sub-goals in order to assist the achievement of main goals.

Here, we already defined organization requirement as

$$Org_{requirement} : \left(Org_{objective}\right)\left(Org_{actions}\right) \tag{13}$$

and organization process, as Eq. (14), and assuming organization requirement as goal and objective as sub-goals

$$Org_{goal} = \left(Sub_{goals}, A_{ctions}\right) \tag{14}$$

So, full relationship can be develop here as

$$\mathrm{Org_{goal}}=\left(\mathrm{Sub_{goals}},\ \mathrm{A_{ctions}}\right),\ \mathrm{Org_{process}}=\left\{\mathrm{Org_{process}(IN,\ OT)}|\ \mathrm{IN}\ \in\ \mathrm{P_i},\ \mathrm{OT}\ \in\ \mathrm{P_o}\right\}. \tag{17}$$

Summarized as,

$$\mathrm{Org_{goal}} = \left\{\mathrm{Org_{goal}}(x,\ y)|x\ \in\ \mathrm{A_{ctions}}, y\ \in\ \ \mathrm{Sub_{goals}}\right\} \tag{15}$$

So,

$$\mathrm{Org_{goal}} = \left\{\mathrm{Org_{goal}(IN,\ OT)}|\ \mathrm{IN}\ \in\ \mathrm{A_{ctions}},\ \mathrm{OT}\ \in\ \mathrm{Sub_{goals}}\right\}. \tag{16}$$

It summarizes the entire relationship as Org_{goal} rely on Sub_{goals} and A_{ctions}. But as we understand, organization relies on data to support achieving the goals. Therefore, full relationship is defined as Sub_{goals} is an *OT* of *Act* where Sub_{goals} and A_{ctions} are the requirement for Org_{goal}. Ontology has a set of organization goal OG elements.

$$\text{Ontology (O)} = \text{Organizational goals (OG)} \tag{17}$$

a set of sub-goal SG, action A and task T relation on the elements of OG, respectively.

For 2 ontologies O_1 and O_2, the set of possible pair of element is defined as

$$\mathrm{OG} = (\mathrm{SG_1} \times\ \mathrm{SG_2}) \cup (\mathrm{A_1} \times\ \mathrm{A_2}) \cup (\mathrm{T_1} \times\ \mathrm{T_2}) \tag{18}$$

with SG_1, A_1 and T_1, the sets of elements of OG in O_1, and SG_2, A_2 and T_2, the sets of elements of OG in O_2.

Then, denote pair of elements (e_1, e_2) refers to the same elements. In other words, each elements $e_1 \equiv SG_1 \cup A_1 \cup T_1$, corresponds to on elements $e_2 \equiv SG_2 \cup A_2 \cup T_2$, and vice versa.

On the other issue, some issue may occur with the OG elements. "A" may have sub-action SA, and "T" may have a sub-task ST. Thus,

$$SG = (A_1 \times A_2) \cup (SA_1 \times SA_2) \tag{19}$$

$$A = (T_1 \times T_2) \cup (ST_1 \times ST_2) \tag{20}$$

with A_1 and SA_1, the set of element of SG_1, and A_2 and SA_2, the set of element of SG_2. T_1 and ST_2, the set of element of A_1, and T_2 and ST_2, the set of element of A_2. The elements of the relationship implies as:

Organization element
OE ≡ {Set of all possible organization elements}
G ≡ {Set of goal}
SG ≡ {Set of sub-goal}
A ≡ {Set of action}
T ≡ {Set of task}

$$OE \equiv G \cup SG \cup A \cup T \tag{21}$$

Variables
V ≡ {Set of all possible variables}
V_x ≡ {Set of all dependent variables}
V_y ≡ {Set of all independent variables}

$$V : x = (n) \text{ with } n \in V_x \tag{22}$$

$$V : y = (n) \text{ with } n \in V_y \tag{23}$$

Relationship

R ≡ {Set of possible relationship}
OR ≡ {Set of organization relationship}
VR ≡ {Set of variables relationship}
R ≡ f (OE(G(SG(A(T)))))
OE ≡ f (G(SG(A(T))))
G ≡ f (SG(A(T))))
SG ≡ f (A(T)))
A ≡ f (T)

Define as

$$x \equiv f(y) \tag{24}$$

Therefore, we show the entire relationship as

$$OE \equiv f(G(SG(A(T)))) \tag{25}$$

where OE is a relationship between G, SG, A and T. In order to look at the possible variable, we define the relationship based on several factor where sum of $\sum$ and number of size n. The factors are:

if sub-goal implies
cause goal

$$\frac{SG(n)}{(\sum SG)} = G \tag{26}$$

if action implies
cause sub-goal

$$\frac{A(n)}{(\sum A)} = SG \tag{27}$$

if task implies
cause action

$$\frac{T(n)}{(\sum T)} = A \tag{28}$$

Based on the factor definition, we defined the variable as

$$x \equiv f\,(y) \tag{29}$$

where, variable x depend on variable y. Based on the variables, we define G as

$$\mathrm{G} \equiv f\,(y) \tag{30}$$

$$\mathrm{G} \equiv f\,(\mathrm{SG}) \tag{31}$$

$$\mathrm{G} \equiv f\,(\mathrm{SG_i} \wedge \mathrm{SG_n}) \tag{32}$$

where, G depend several number of SG. For example SG1 and SG2, SG2 and SG3. Thus, we assume G depend on the number of SG as written below

$$G = \frac{\sum G}{n} \tag{33}$$

$$G = \frac{SG(n)}{(\sum SG)} \tag{34}$$

If we define SG, then

$$\mathrm{SG} \equiv f\,(y) \tag{35}$$

$$\mathrm{SG} \equiv f\,(\mathrm{A}) \tag{36}$$

$$\mathrm{SG} \equiv f\,(\mathrm{A_i} \wedge\ \mathrm{A_n}) \tag{37}$$

where, SG depend several number of A. For example A1 and A2, A2 and A3. Thus, we assume SG depend on the number of A as written below

$$SG = \frac{\sum SG}{n} \tag{39}$$

$$SG = \frac{A(n)}{(\sum A)} \tag{40}$$

If we define A, then

$$\mathrm{A} \equiv f\,(y) \tag{41}$$

$$\mathrm{A} \equiv f\,(\mathrm{T}) \tag{42}$$

$$\mathrm{A} \equiv f\,(\mathrm{T_i} \wedge \mathrm{T_n}) \tag{43}$$

Where, A depend several number T. For example T1 and T2, T2 and T3. Thus, we assume A depend on the number of T as written below

$$A = \frac{\sum A}{n} \tag{44}$$

$$A = \frac{T(n)}{(\sum T)} \tag{45}$$

The other situation, if organization only require one of the variable between SG and A, then we define the relationship as

$$\mathrm{G} \equiv f\,(\mathrm{SG_i} \vee\ \mathrm{SG_n} |\ \mathrm{SG_i} \supset \mathrm{SG_n}) \tag{46}$$

where, G rely on several number of SG. For example, G relies on SG1 or SG2, where SG2 depend on SG1. We define SG as

$$\mathrm{SG} \equiv f\,(\mathrm{A_i} \vee \mathrm{A_n} |\ \mathrm{A_i} \supset \mathrm{A_n}) \tag{47}$$

where, SG rely on several number of A. For example, SG relies on A1 or A2, where A2 depend on A1. Then we define A as

$$\mathrm{A} \equiv f\,(\mathrm{T_i} \vee\ \mathrm{T_n} |\ \mathrm{T_i} \supset \mathrm{T_n}) \tag{48}$$

where, A rely on several number of T. For example, A relies on T1 or T2, where T2 depend on T1.

4 Metric Requirement

We define metric requirement as a metric design of what needs to be accomplished during the metrics process. Specific team attempts to identify the needs toward organization objective. Thus, we identify two variables toward metric requirements which are verifiable and measure. Verifiable: a set of data that been agreed for converting process into measure. Measure: characteristics in a numerical or nominal form.

4.1 Metric Analysis

Metric analysis is defined as a requirement that must be fulfilled toward metric development. We identify three variables toward metric analysis which are control, communication and improvement.

Control: The ability of metric to evaluate and control the source that they measure.
Communication: The ability of metric to communicate externally and internally for the purpose of control.
Improvement: The ability to identify the gaps for improvement.

Based from the discussion of metric requirement and metric analysis, the metric can be written as

Metric: (MetReq)(MetAna)

and

Metric: $(V_{erifiable}, M_{easure})$ $(C_{ontrol}, C_{ommunication}, I_{mprovement})$

where, the metric requirement and metric analysis are the variables of metric model. Verifiable and measure are the characteristics for metric requirement. Control, communication and improvement are the characteristics for metric analysis. Metric model in this paper allow the evaluation of organization data toward gaps, setting and change. We denote sub-goal as SG, action as A, sub-action as SA, task as T and sub-task as ST for our proposition example. We assume that

$$\begin{aligned} &SG \rightarrow (A \rightarrow SG)\ (SA \rightarrow A) \\ &\qquad (T \rightarrow A)\ (ST \rightarrow T) \text{ and} \\ &(SG \rightarrow \neg A) \rightarrow (SA\ (SG \rightarrow \neg A) \end{aligned}$$

SG is implied if $\neg A$ is true and SA is implies to achieve SG and A
$(A \rightarrow \neg SA) \rightarrow (T\ (A \rightarrow \neg SA)$
A is implied if $\neg SA$ is true and T is implies to achieve A and SA
$(SA \rightarrow \neg T) \rightarrow (ST\ (SA \rightarrow \neg T) \rightarrow (T \rightarrow \neg ST)$
SA is implied if $\neg T$ is true and ST is implies to achieve SA and T. T is implies if $\neg ST$ is true.

5 Case Study

5.1 The World Bank

Most data in the World Bank dataset comes from the governments of individual countries. The World Bank collects data on living standards and debt, but not much else. Some also comes from various international and national agencies with which World Bank partners (see http://ucatlas.ucsc.edu/data.html). These data are grown from the commitment of the member countries that participate with the development projects since early 1950s. The variety of data types is also diverse and data are not just created in spreadsheets or tables but also in charts, maps and audio/video. Therefore, managing and analyzing these data can be very challenging.

The World Bank collects and processes large amounts of data and generates them on the basis of economic models. These data have gradually been made available to the public in a way that encourages reuse. The World Bank stores data of reconstruction and development for 188 countries on 20 different topics. The topics incorporate different data indicators that store data from 1960 to 2014. For example, one of the data indicators for economic growth is agriculture (value added). The data are collected from 1960–2014 for 188 countries. This dataset create large amount of data only for one indicator under one topic.

The World Bank provides an analysis and visualisation tool that contains collections of time series data on a variety of topics that allow us to create our own queries. Therefore, we can generate tables and dashboards as new knowledge to be shared. In this case study, we present queries as the goals. Then we capture relevant data from huge amount of the datasets in relation to the goals. We define the metrics to analyse this data and present it in the dashboard to support decision-making in light to the goals.

5.2 An Ontology for the World Bank

In this paper, the queries are created from the topics. We define these queries as goals. In order to evaluate these topics, we develop an ontology for the World Bank to filter large collection of data. Therefore, we can identify the relationship between the topics and data indicators. An ontology creates knowledge to helps us to define the goals that we want to evaluate. We can make a decision on which goals we want to evaluate and which data are relevant to the goals.

The amount of data stores in the World Bank make it difficult for us to identify the goals that we want to evaluate. Based on the World Bank's website, data can be identified from the topics and data indicators. Figure 2 shows the relationship based on an ontology that classifies the topics.

Based on these topics, there are few data indicators that can be identified. Figure 3 shows some examples of data indicators that relate to the certain topics. This ontology helps us to create any query.

In this paper, we refer ourselves as information professional. Using the information that we have, we set the goals that we want to evaluate. Based on the goals, we can capture relevant data that relate to the goals.



Fig. 2. Data are extracting the topics from the World Bank using Protégé OntoGraf.

5.3 Identify the Goals for the World Bank

In this case study, we aim to analyse the level of the economic growth in South East Asia in 2013. We decide to analyse the development in Indonesia, Cambodia, Malaysia, Philippines and Singapore. According to the World Bank, economic growth is central to economic development. When national income grows, real people benefit. Data can help policy-makers better understand their countries' economic situations and guide any work toward improvement. Data here covers measures of economic growth that includes indicators representing factors known to be relevant to economic growth, such as industry, manufacture, services, agriculture and household.

In order to achieve this aim, we analyse these five indicators that we believe is important to evaluate the development level of the economic growth in South East Asia. This is how we want to define the goals in this paper. We are mindful of the fact that information professionals might define the goals in a different way to the way we



Fig. 3. Data are extract using Protégé OntoGraf to shows the indicators from four different topics.

have undertaken to define the goal, which would require a different approach to evaluate the goals.

Economic growth is central to economic development. When national income grows, real people benefit. While there is no known formula for stimulating economic growth, data can help information professionals to assist the policy-makers with better understand the economic situations in South East Asia and guide any work toward improvement. Data here covers measures of economic growth based on the value added for industry, agricultural, manufacturing, service and household.

According to the World Bank website, value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The origin of value added is determined by the International Standard Industrial Classification (ISIC), revision 3. In this case study, data are in current U.S. dollars.

- Five goals in relation to the economic growth, as shown in Fig. 3.

Goal 1: Development of services for wholesale and retail trade

Services correspond to ISIC divisions 50–99. They include value added in wholesale and retail trade (including hotels and restaurants), transport, and government, financial, professional, and personal services such as education, health care, and real estate services. Also included are imputed bank service charges, import duties, and any statistical discrepancies noted by national compilers as well as discrepancies arising from rescaling.

Goal 2: Development of agricultural production

Agriculture corresponds to ISIC divisions 1–5 and includes forestry, hunting, and fishing, as well as cultivation of crops and livestock production.

Goal 3: Development of manufacturing to supports industries growth

Manufacturing refers to industries belonging to ISIC divisions 15–37.

Goal 4: Development of industry to supports sustainable growth

Industry corresponds to ISIC divisions 10–45 and includes manufacturing (ISIC divisions 15–37). It comprises value added in mining, manufacturing (also reported as a separate subgroup), construction, electricity, water, and gas.

Goal 5: Development of household value for goods and services

Household is the market value of all goods and services, including durable products (such as cars, washing machines, and home computers), purchased by households. It

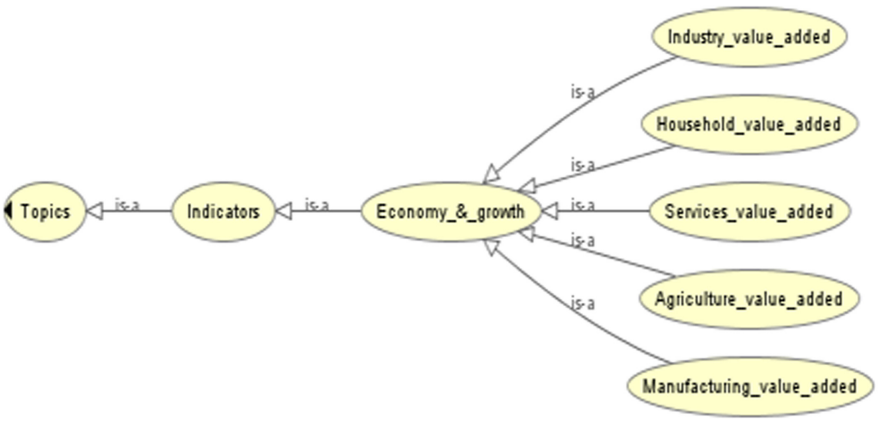

Fig. 4. Protégé OntoGraf for the World Bank to evaluate the development of economic growth in South East Asia.

excludes purchases of dwellings but includes imputed rent for owner-occupied dwellings (Fig. 4).

6 Conclusions

The discussion include how the organizational goals are identified, the possible sub-goals and variables that relate to the organizational goals, how to identify data that has been presented to the organizational goals, how to analyze organizational data and use this analysis results as a feedback to evaluate the level of the organizational goals achievement.

This paper discussed on how the framework aims to be applicable, repeatable, flexible and configurable in many domains. Big data analytics could be used to examine large amounts of data of a variety of types to discover useful information. Such information can provide a competitive advantage to help better decisions be made in relation to the organizational goals. The primary goal of big data analytics is to help companies make better business decisions by enabling data analyst to analyze huge volumes of transaction data which remains an unresolved challenge for conventional business intelligence programs.

The challenge in terms of big data analytics includes how the framework searches, transfers, captures and analyze the data from the vast collection of datasets. Due to the flexibility of the framework, it can capture large sets of data that relate to the organizational goals. This data will be analyzed and the results can assist the decision-making process in relation to the organizational goals.

References

1. Ordonez, C., Zhang, Y.Q., Cabrera, W.: The gamma matrix to summarize dense and sparse data sets for big data analytics. IEEE Trans. Knowl. Data Eng. **28**, 1905–1918 (2016)
2. Ortega, F.M., Subias, J., Cassany, D.: An ethnographic approach to digital literacy in a Compulsory Secondary Education Institute in Barcelona. CPU-E Revista De Investigacion Educativa, pp. 190–215, July–December 2016
3. Papakonstantinou, P.A., Woodruff, D.P., Yang, G.: True Randomness from big data. Sci. Rep. **6**, 33740 (2016)
4. Park, E.-M., Seo, J.-H., Ko, M.-H.: The effects of leadership by types of soccer instruction on big data analysis. Cluster Comput. **19**(3), 1647–1658 (2016). https://doi.org/10.1007/s10586-016-0609-2
5. Phuapan, P., Viriyavejakul, C., Pimdee, P.: An analysis of digital literacy skills among Thai University Seniors. Int. J. Emerg. Technol. Learn. **11**, 24–31 (2016)
6. Kobayashi, K., Kaito, K.: Big data-based deterioration prediction models and infrastructure management: towards assetmetrics. Struct. Infrastruct. Eng. **13**, 84–93 (2017)
7. Broderick, S.R., Santhanam, G.R., Rajan, K.: Harnessing the big data paradigm for ICME: shifting from materials selection to materials enabled design. JOM **68**(8), 2109–2115 (2016). https://doi.org/10.1007/s11837-016-2019-6
8. Puthal, D., Nepal, S., Ranjan, R., Chen, J.J.: A dynamic prime number based efficient security mechanism for big sensing data streams. J. Comput. Syst. Sci. **83**, 22–42 (2017)

9. Wikibon. Big data vendor revenue and market forecast 2013–2017, 2014, 26 June 2015. http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017
10. Sharma, S., Osei-Bryson, K.-M.: Organization-ontology based framework for implementing the business understanding phase of data mining projects. In: International Conference on System Sciences, Hawaii, p. 27 (2008)
11. Izhar, T.A.T., Torabi, T., Bhatti, M.I., Liu, F.: Recent developments in the organization goals conformance using ontology. Expert Syst. Appl. **40**, 4252–4267 (2013)
12. Schoeneborn, D., Kuhn, T.R., Kärreman, D.: The communicative constitution of organization, organizing, and organizationality. Organ. Stud. **40**, 475–496 (2019)
13. Wu, B.: The semantic retrieval system for learning resources based on subject knowledge ontology. Adv. Comput. Sci. Res. **80**, 467–469 (2018)