Early Information Requirements Engineering for Target Driven Data Warehouse Development

Naveen Prakash and Hanu Bhardwaj

MRCE, Sector 43, Surajkund Badhkal Road, Faridabad 121001, India
praknav@hotmail.com, bhardwajhanu@gmail.com

Abstract. We propose two stages for data warehouse requirements engineering (i) an 'early information' part where the information relevant to decision making is discovered, and (ii) a 'late' part where this information is structured as facts and dimensions. Our focus is on the former. Early information data warehouse requirements engineering starts with targets defined as pairs of the form <A, I> where A is an aspect of an organization and I is a set of business indicators. An aspect is a work area, work unit, service or quality to be preserved in an organization. Business indicators are measures/metrics for **specifying** the desired performance level of aspects. Targets are organized in a target hierarchy. This hierarchy is a complete specification of what is to be achieved by a top level target. We associate targets with choice sets so that alternative ways of target achievement can be represented. These alternatives form their own hierarchy. Finally, information relevant to selection of each alternative is discovered through Ends, Means, Key Success Factor, and Outcome Feedback analysis techniques. These techniques determine early information that is to be subsequently to be processed in the 'late information' requirements engineering stage. Our early information requirements engineering phase is illustrated through a case study.

Keywords: aspects, business indicators, target, choice set, early information.

1 Introduction

Starting from its early years when requirements engineering of data ware house systems was about the last thing to be handled [Inm96], requirements engineering for data warehousing has acquired increasing significance. Three life cycles for DW development were presented in [Pra08], namely, database schema driven, ER diagram driven, and Goal driven respectively. Goal oriented approaches [Boe99, Boe00, Bon01, Gol99, Pra04, Pra08] consider organizational and stakeholder goals to define DW structure. Horkoff and Yu [Hor12] observe that "goal modeling is not yet widely used in practice" but has been successfully applied in case studies. The inhibiting factors in adoption of goal oriented techniques were identified in [Rol03] as (a) Domain experts find the notion of a goal to be fuzzy which makes it difficult to deal with it, and (b) Goal reduction is not a straight forward process. Evidently, these issues get carried over to data warehouse requirements engineering as well.

K. Sandkuhl, U. Seigerroth, and J. Stirna (Eds.): PoEM 2012, LNBIP 134, pp. 188–202, 2012.

Recall that the data warehouse requirements engineering problem is to define the information contents of a data warehouse to support the decision making task at hand. We propose to partition this task into (i) an 'early information' part where the information relevant to decision making is discovered, and (ii) a 'late' part where the discovered information is structured as facts and dimensions. Our focus is on the former.

Early information requirements engineering starts off by determining targets that organizations want to achieve. A target is defined as a pair, <A, I> where A is an organizational aspect and I is a set of business indicators. An aspect represents a work area, work unit, service or quality in an organization. Business indicators are measures/metrics of aspect performance and can be discovered in organizations using techniques like Balance Scorecard. As an example of a target consider sales. Sales is an aspect of an organization. It is a service. An appropriate business indicator might be 'Number of units sold should increase by 5%'. The notions of aspects and business indicators provides a relatively clear basis for defining targets as compared to goals in goal orientation.

Target reduction can be done by following the aspect driven or indicator driven approach. In the former, sub aspects of a top level target are determined and their business indicators are found to form sub targets. In the latter, sub business indicators that go into computing the top level business indicators are determined, their aspects found and sub targets formulated. Thus, we have a relatively clear basis for guiding target reduction as compared to goal reduction.

Now we can consider the second step of our 'early information' requirements engineering stage. Whereas target hierarchies identify what the organization wants to achieve, decision making is to be done to indeed achieve targets. To reflect this, we associate choice sets with targets. We define a choice set as a pair <CSO, O> where CSO is the choice set objective and O is the set of alternatives for meeting the CSO. Analogous to target reduction, we have choice set reduction. Each sub choice set is associated with the target that it affects.

At this moment, the full decisional capability required for target achievement is known. Now, in the last stage of early information requirements engineering, we discover the information required to make the most suitable selections from choice sets. By 'early' we mean that information is determined in an abstract, relatively fuzzy form devoid of any structure. Yet, all requirements, for example, of history and aggregation shall be identified here. In this paper, we have relied on the tool reported in [Pra09] for obtaining this early information. Our approach for the 'late information' stage can be found in [Pra08].

Let us compare the early information with the notion of early requirements. Early requirements engineering [Yu 97] is concerned with modeling and analyzing the operational environment in which the software system shall eventually function. It involves an examination of the objectives, business processes, and interdependencies of different stakeholders. Our early information is arrived at by considering the decisional environment that the data warehouse shall support. However, it is nor 'early requirements' because inter dependencies between stakeholders are not investigated. We place 'early information' as the first stage of 'late requirements' engineering for data warehouses where a first cut identification of information is obtained.

The approach of this paper has been applied to a case study of a Dairy Development Board. The paper uses examples from this to illustrate the various concepts. A part of this case study is then presented.

The layout of the paper is as follows. In the next section we consider our notion of targets. It explains the two dimensions of a target and shows how choice sets can be constructed. Section 3 presents the movement from alternatives expressed in choice sets to the required information. Section 4 contains the case study to illustrate the applicability of our proposals. Section 5 places our proposals in existing literature and section 6 is the concluding section.

2 Targets and Their Dimensions

We define a target as a pair <A, I> where A is an aspect of an organization and I is the set of indicators for A. An aspect is a work area, a work unit, a service, or a quality to be preserved. Consider an animal husbandry department. Examples of aspects in this department are as follows:

1. Genetics and Breeding are two work areas in the animal husbandry department.
2. Breed Conversion and Reproduction are two work units.
3. Veterinary Facility and Extension Programme are services provided by the department.
4. Distributed Services Across Regions is a quality to be reserved by the department.

A business indicator is a metric whose value identifies 'what is to be achieved'. This value may be absolute (27) or it may be relative (current +5%). A business indicator may be simple or composite. A composite business indicator is one whose value is obtained from its component business indicators. The notion of a business indicator is similar to that of a key performance indicator KPI. It is explained in [Bar10] that a KPI is a metric associated with goals. Thus percentage increase in sales is a KPI for the goal Increase market share. A business indicator, I, on the other hand, is associated with an aspect A to form a target. There is thus a conceptual difference between a business indicator and a KPI, the former is a **constituent** of a target whereas a KPI is external to a goal. Further, a business indicator is not associated with a goal but with an aspect which is not a goal but a work area, unit etc. as pointed out above.

In forming a target, we create an association between aspects and business indicators. A target is thus able to specify 'what is to be achieved' by the aspect. Notice that a target may have an aspect that is associated with more than one business indicator. For example, the aspect Breed conversion may have two business indicators, (current insemination + 10%) and (hybrid animals + 5%).

The notion of a target is depicted in Fig. 1. The pair <A, I> is shown by the two conjoined boxes. The arrow from A to I shows the indicators for A. The self loop on I shown by the solid arrow shows the relationship between a composite indicator and its component indicators. The arrow from I to A shows that aspects can be associated with indicators. Thus there is a bi-directional relationship between aspects and indicators.

The bidirectional A-I relationship allows us to form a target hierarchy. Let us be given a top level target. There are two approaches to target reduction.

1. Aspect driven reduction. Here, sub aspects of the top level aspect are determined. For example, if the top level aspect is a work area, then the work units comprising the work area are determined. Similarly, for a work unit, the services offered and the qualities to be preserved form its sub units. Appropriate business indicators are associated with the sub aspects to form sub targets.
2. Indicator driven reduction. If the top level business indicator is composite, then the components that go into computing its values are determined. The aspects responsible for these are associated with the sub indicators and values to be achieved are identified through brain storming sessions. Thus, we formulate sub targets of top level targets.

The notion of a target identifies what is to be achieved by different aspects of an organization. It is a prescription and provides purpose for the decision making to be carried out in the organization. Decision making requires a choice set for selection of an alternative that is most helpful in achieving the target. Thus, we need to associate choice sets with targets, and formulate alternatives. This is modeled in Fig. 1 through the fulfils relationship between targets and choice sets.

The right side of the figure shows that a choice set is a pair <CSO, O>. CSO is a Choice Set Objective that can be met through different options, O. As shown, the relationship, fulfils, identifies which target is affected by which choice set. This 'affectation' occurs because of the capacity of options to cause changes in the values of indicators. The self loop on O, labeled 'achieves', shows that options can have their own component options. This gives rise to the achievement hierarchy.

As for target hierarchies, it is possible to build hierarchies of choice sets. Again there can be two approaches, CSO driven and options driven that work in a manner analogous to the aspect and indicator driven approaches considered above.

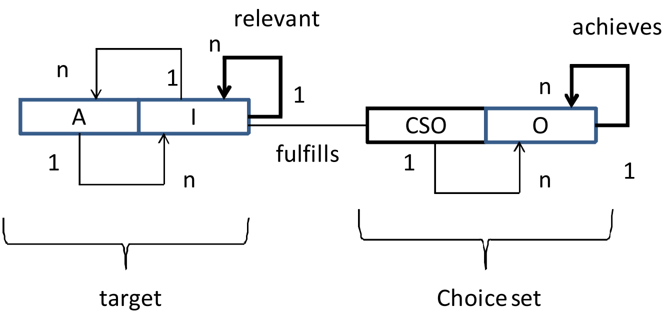

Fig. 1. A Conceptual Overview

We view Fig. 1 as giving us a two dimensional framework for targets. The first dimension is called the relevance dimension and corresponds to the left part of the figure. In this dimension a hierarchy is set up using the self loop whose nodes are targets and edges specify a 'relevance' relationship between nodes. The second dimension is called the fulfillment dimension and corresponds to the right hand side of the figure. It models the different ways in which an indicator can be affected and comprises of the hierarchy corresponding to the self loop on O.

2.1 The Relevance Dimension

The starting point of the determination of the information to be kept in a data warehouse is found in the relevance dimension. Targets are organized in a 'relevance' hierarchy, defined as follows:

Relevant(t,T) iff di_t is a factor in estimating di_T.

Here, di_T refers to a desired indicator of T and di_t refers to a desired indicator of t. We say that t is a sub target of T. If there is no t such that Relevant(t,T) then T is an atomic target and is at the leaf of the relevance hierarchy.

Consider the meaning of estimation. Estimation can range from completely well defined estimation to purely judgmental. We illustrate this by considering a target T and its relevant target t. Consider Di_T as expressed in the set of indicators { A_1, A_2, ….. A_n} and Di_t in {b_1, b_2, …. , b_m}. Then estimation can take on any of the forms as follows:

1. Identity estimation: in this case $A_i = b_j$. For example let $Di_{\text{Control Population}}$ be defined as {number, birth rate, death rate, female_male_ratio}. Let $Di_{\text{Birth Rate Control}}$ = {birth rate, calf mortality rate}. Here, 'Birth rate' of 'Control Population' is equal to 'Birth rate' of 'Birth Rate Control'.
2. Derived estimation: Ai = F(bj). Again taking 'Birth rate' as example, it is possible that interest in 'Control Population' is in the Net birth rate taken as the difference between 'Birth rate' and 'Calf mortality rate' of 'Birth Rate Control', i.e., Birth rate $_{\text{Control Population}}$ = Birth Rate $_{\text{Birth Rate Control}}$ – Calf Mortality Rate $_{\text{Birth rate Control}}$. Functions like max, min etc. may also be applied.
3. Ill-structured estimation: In this case, one may use techniques like trend analysis, forecasting, extrapolation etc. to estimate Ai from bj. Perhaps, there is a trend captured in the factor, 'Population elasticity' that helps us to estimate 'New population' based on 'New birth rate' and 'Old population'. Then using this, estimation of 'Control Population' can be made. The proposals of [Bar11] are of interest here.
4. Judgmental estimation: This is the case where estimation is completely ad-hoc and undefined. It is a pure judgment on the part of the decision maker. A rich ground for judgmental estimation is ncc targets.

2.2 The Fulfillment Dimension

Whereas the relevance dimension is concerned with targets and their relevance to one another, the fulfillment dimension deals with the alternative ways for causing the change in indicator value.

2.2.1 Choice Set

We view a choice set as an abstraction, a class of elements that have the same purpose, that is, a choice set must be coherent. An element is a member of a choice set **if and only if** it meets the coherence property.

Coherence says that all elements of a choice set must achieve the same purpose. As pointed out earlier, this purpose is captures in the notion of a choice set objective, CSO. For example, consider the choice set, CSET = {Improve cattle health, Ensure cattle nutrition, Improve animal husbandry} having the CSO, 'Improve milk production'. All elements of this set have the same objective to 'Improve milk production'. CSET is coherent. As an example of an incoherent choice set consider $CSET_1$= {Improve cattle health, Ensure cattle nutrition, Ensure budgetary compliance}. The element, 'Ensure budgetary compliance' does not contribute to the stated CSO. Therefore, $CSET_1$ is not coherent.

For a coherent choice set, we may have members that are CSOs and/or functions. We illustrate this by considering the choice set, 'Improve milk production', in its different forms as follows:

i. The set {Improve cattle health, Ensure cattle nutrition, Improve animal husbandry} has all its elements that specify choice set objectives.
ii. Consider the choice set for 'Improve cattle health', {Perform de-worming, Perform brucella vaccination, Perform mastitis vaccination}. All options of this choice set are functions to be carried out.
iii. For the CSO 'Ensure cattle nutrition', let us be given the choice set {Launch nutrition education programme, Maintain quality feed stock, Improve raw material procurement}. Here 'Launch nutrition education programme' is a CSO whereas 'Maintain quality feed stock' is a function.

Thus, it is possible to build a choice set hierarchy for every CSO. This is done by determining if a CSO/function provides an alternative way of meeting another CSO. We capture this in the 'achieves' relationship defined as follows:

Achieves(CSO_i,CSO) iff CSO_i is an alternative way of meeting CSO,

OR

Achieves(F_i,CSO) iff F_i is a function and is an alternative way of meeting CSO.

2.2.2 Fulfillment

The fulfillment dimension associates appropriate choice sets with targets of the relevance hierarchy. As mentioned earlier, a target may have more than one indicator, for example, $Di_{Control\ Population}$= {number, birth rate, death rate, female_male_ratio}. Each indicator has its own choice set.

We define appropriateness as the alignment of the choice set to its indicator. That is, a choice set is **appropriate** to a target if it identifies ways to change an indicator of the target. For example, the coherent choice set CSET of section 2.2.1 above is appropriate to the indicator 'cattle milk production cmp, cmp + 4%' of the target 'Milk Production'.

We can now define the fulfillment criterion as follows:

Fulfils(CSO, I_m) iff CSO is appropriate to $I_{m,}$ where I_m refers to a indicator of T.

Fulfils says that the elements of CS individually or jointly cause the indicator of T to be reached.

3 Determining Early Information

As mentioned earlier, we determine 'early' information to be kept in the warehouse. This early information is unstructured and at an abstract level. First, notice that indicators directly identify the needed information, for example, birth rate, death rate, and can be obtained from the relevance hierarchy. The interesting question is to determine information for selecting an alternative. We refer to this as pertinent information.

In [Pra09] we handled the issue of pertinent information through a tool called Raju that elicits this information. The tool uses four elicitation techniques called ends analysis, means analysis, key success factors, and outcome feedback analysis. For each element of a choice set, the requirements engineer applies these four techniques to elicit what that paper calls parameters. These parameters are then analyzed to determine which of these parameters themselves lead to choice sets and which identify information. This is done on the basis of heuristics. Pertinent information is obtained as

1. History to be maintained: for example, information of registered members must be available for the last 5 years.
2. Category-wise information: for example, month-wise, age wise, designation-wise etc.
3. A report or document.
4. A computation obtained by applying a function like Count of, Average etc.
5. Comparison information: for example, comparison of our performance against others.

Pertinent information is elicited for all members of all choice sets.

4 Case Study

The case study considered in our paper is of a Dairy Development Board, DDB. In the current analytic environment at DDB, there is lack of access to consistent and integrated decision making data. The current reporting system and decision support environment is not addressing the needs of decision makers across the DDB, which causes deficiencies in data ownership and data access policies. Thus requirement of DDB is to anticipate the future and to identify data for decision making, so that it prepares itself for the emerging competitive environment and provides developments in quality of milk and genetics of animals. The information required by DDB is for strategic purposes such as trend identification, forecasting, competitive analysis and targeted market research, so as to ensure a better future for farmers and milk producers.

The activities of DDB are linked to Dairy Cooperative Societies DCS at village level, Milk Producer Union at district level, Milk Marketing federations at state level, Cattle feed plants, and research institutions, and can be broadly divided into three aspects, namely, financial, technical, and legal.

Here we apply our target oriented approach to DDB. We develop the relevance and fulfillment dimensions, choice sets and pertinent information. For reasons of space, we shall illustrate our ideas with a part of the total case.

The top level targets are shown in Table 1. The indicators and aspects of the targets were determined by the Board through a process of discussion and brainstorming.

Table 1. The three targets of Dairy Development Board

Sr. No.	**Target**		**Sub Target**	
	Aspect	**Indicator**	**Aspect**	**Indicator**
1	Finance support	Budget Compliance	Fund Management	<[current recovery cr, cr+10%], [current disbursement cd, cd +3%]>
			Loan Management	<current backlog cb, cb – 20%>
2	Technical support	Efficiency	Milk Production	<[cattle milk production cmp, cmp + 4%],[cattle in milk cam, cam + 60%]>
		Effectiveness	Milk Procurement& Distribution	<[dcs satisfaction percent dcssp, dcssp + 10%]>
3	Legal matters	DCS coverage	Legal framework developed	<current DCS cdcs, cdcs+3>
			Adequately staffed legal cell	<[dcs to legal officer ration dlo, dlo= 20:1]>

In the relevance dimension of Technical support shown in the table, the indicators of sub aspect, ‘Milk production’, can be quantified: new cattle milk production must be current cattle milk production, cmp, plus 4%. Similarly, cattle in milk must rise by 60% i.e. cam must become cam+60%.

Further, the sub-aspect ‘Milk procurement and distribution’ has the indicator quantified as <[dcs satisfaction percent dcssp, dcssp + 10%]>

Now consider the target about ‘Milk production’. Its relevant sub target hierarchy is shown in Fig. 2. An explanation of the sub aspects is provided in the table below:

Target	**Sub Aspect**	**Scope of Sub Aspect**
Milk Production, , <[cattle milk production cmp, cmp + 4%],[cattle in milk cam, cam + 60%]>’	Genetics and breeding service	Genetic Improvement, Breed conversion
	Animal reproduction	Production of good quality semen and creation of artificial insemination facilities
	Fodder and feed service	Production and distribution of fodder and feed
	Animal health	Disease diagnostic and control

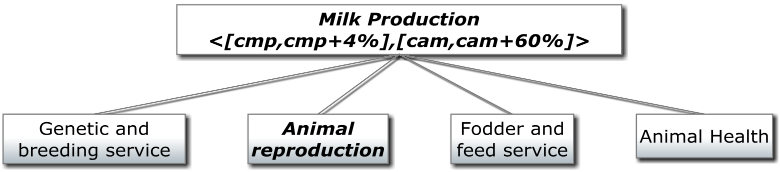

Fig. 2. The Relevant sub aspects of target about ‘milk production’

Now consider **aspect** ‘Animal reproduction’; its sub targets are shown in Fig. 3. The indicators of its sub aspects are ‘increasing the semen stations’, st, to 11 from the current st; ‘raising artificial insemination centres’, aic, from 20,000 to 38,000; ‘raising field artificial insemination’, fai, from 4.20 min to 10.30 min, and ‘increasing semen production’, sp, from 6.20 min to 13.20 min.

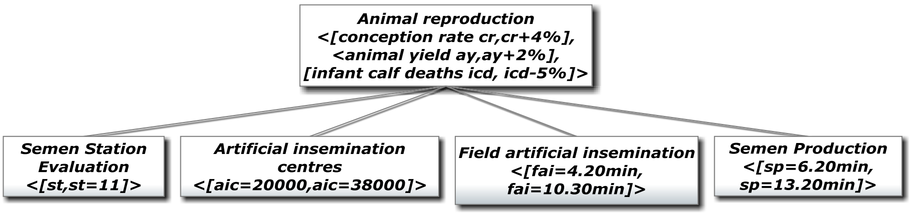

Fig. 3. The Relevant sub targets of target ‘Animal reproduction, <[conception rate cr,cr+4%], <animal yield ay,ay+2%], [infant calf deaths icd, icd-5%]>

There are no sub targets of any of the leaf targets of Fig. 3. It can be seen that Fig. 1 to 3 can be put together to form a three level relevance hierarchy.

Now, **consider the fulfillment dimension** for the target ‘Milk production, <[cattle milk production cmp, cmp + 4%],[cattle in milk cam, cam + 60%]>’’ (Table 1). There are two indicators associated with it, and we get two CSOs in the fulfillment dimension. The resulting fulfillment dimension is shown in Fig. 4.

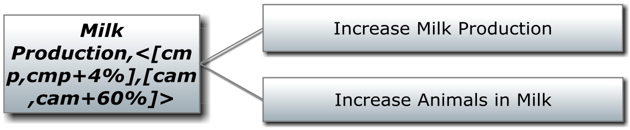

Fig. 4. The Fulfillment Dimension for target ‘Milk Production, <[cattle milk production cmp, cmp + 4%],[cattle in milk cam, cam + 60%]>’

Finally, consider the target for Animal reproduction of Fig. 3. For its three indicators, we get three CSOs as shown in Fig. 5.

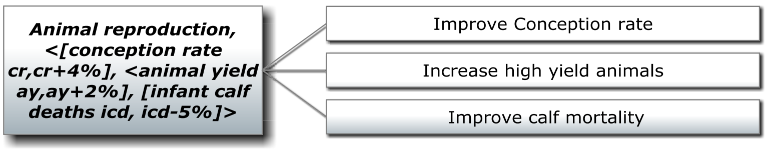

Fig. 5. The Fulfillment Dimension for target ‘Animal reproduction, <[conception rate cr,cr+4%], <animal yield ay,ay+2%], [infant calf deaths icd, icd-5%]>’

Each of the leaf nodes of Fig. 3 has one indicator associated with it; therefore each gives rise to one CSO, as shown in Fig. 6.

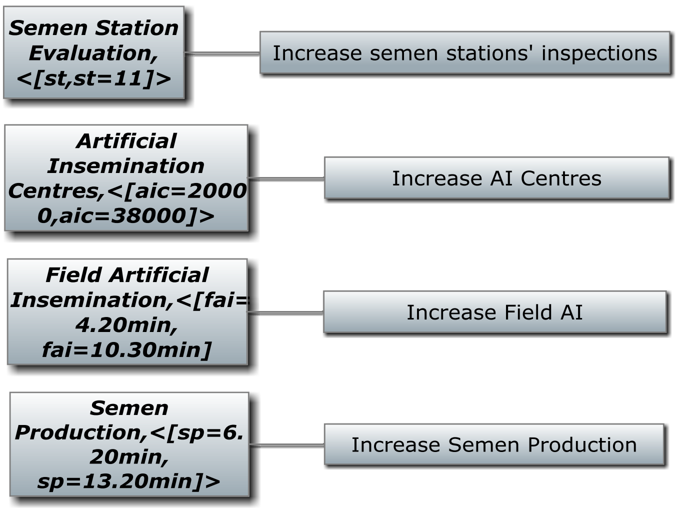

Fig. 6. CSOs of the four leaf nodes

We can now illustrate the Achieves hierarchy. Consider the CSO, 'Increase AI Centres' of Fig. 6. The Achieves relationships, as shown in Fig. 7 are as follows:

a. Achieves(Select location, Increase AI Centres)
b. Achieves(Create AI Franchise Centres, Increase AI Centres)
c. Achieves(Establish proprietary AI centres, Increase AI Centres)

In the foregoing, only 'Select Location' is a CSO whereas the rest are functions. We can further decompose this CSO as follows:

a) Achieves(Select region, Select location)
b) Achieves(Select District, Select Location)

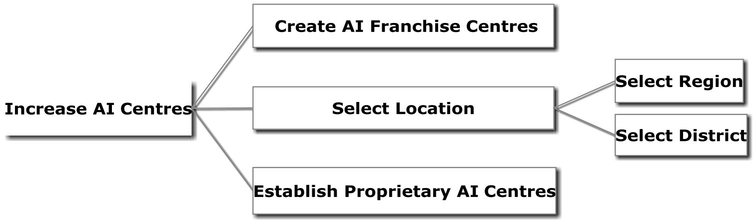

Fig. 7. Achieves Hierarchy of 'Increase AI Centres'

As the last step in the requirements engineering process we need **to discover early information** for each alternative of each choice set. The result is a high level, unstructured view of data warehouse contents. We present in Table 2 the pertinent information for each of the CSO corresponding to the Fig 6.

Table 2. Early Information

Choice Set Objective	Pertinent Information
Increase semen stations' inspections	Bull population under station area. Average collection monthly, quarterly annually. Collection wise ranking of stations. inventory
Increase artificial insemination centres	Total number and number of functional centres. Comparison of centre type to be taken monthly, quarterly and yearly as also by milk union, state and region-wise. Total coverage, comparative average.
Increase field artificial insemination	Top 20 unions performing insemination. Top 20 cluster centres. Average number of inseminations carried out by centre type, All these are monthly, quarterly, and yearly as well as by milk union, state and region wise.
Increase semen production	Semen production state and station wise. Female calves born out of insemination. Breedable cattle and buffalo population, cross breaded cattle and buffalo, milk union, village and district wise. Number of fresh insemination, coverage, percentage of total and number of inseminations annually for the last 3 years

4.1 The Process

In this section we trace the requirements engineering process and exemplify it with the process followed in the case study. The various steps comprising the process are as follows:

1. Determine top level aspects. For the Dairy Board, this was accomplished by identifying the main work areas and services to be provided.
2. Determine business indicators. For each top level aspect, the Board laid down the main indicators. For our top level aspects, the indicators were qualitative. The determination of effectiveness, for example, (see Technical Support in Table 1) would be done by questionnaires and feedback sessions. Provision in the data warehouse for this would be necessary, if it is decided to keep detailed information.
3. Formulate top level targets. This is done by associating aspects and their indicators. Thus, our top level target is <Finance support, Budget compliance>.
4. Decompose Targets. The relevance hierarchy is built by decomposing indicators/aspects. The indicator 'Budget compliance' of Table 1 is decomposed and the sub indicators obtained were recovery and disbursement. This is the indicator driven reduction approach. In contrast the target for animal reproduction of Fig. 3 was decomposed using the aspect driven reduction approach. The sub aspects of animal reproduction were found and then sub targets were set up.
5. Identify Choice Set Objective could occur at varying times. It could happen that as soon as a target/subtarget was identified, the choice set objective was identified due to the association between them. At other times, some alternatives of the choice set were first identified and then the CSO was laid down.
6. Build Achieves Hierarchy. This was rather more structured in the case study. Here, it was possible to focus on the decomposition task and identify the components of a choice set.
7. Identify Information. This part was completely guided by the tool reported in [Pra09].

Our observations from the process are as follows. The first is that discovery of top level aspects can be related to the functional areas and units of the DDB. Further, identification of services provided and main tasks performed could be done through brainstorming sessions. Internal management reports can be a good source for determining aspects.

Once the top level aspects were discovered, the analysis of indicators took on a major role. Indicators are found in planning meetings where targets for different functional units are laid down. Often this is on an annual basis though in India it can be done in a five-year plan perspective as well. The targets presented in this paper were obtained in a five-year plan perspective.

5 Discussion

Our proposals here are directly related to determining the information that shall be stored in the data warehouse. The outcome of our engineering step is not the multi dimensional schema but early information that is unstructured. It identifies at a top level the information that is needed and that is to be structured in subsequent stages. This is one major point of difference between our proposals and [Boe99, Boe00, Bon01, Gol99].

The problem of obtaining business intelligence from indicators has been considered in [Bar11]. Here, a technique has been developed for computing composite indicator values from sub indicator values when faced with ill-defined mathematical functions for computing them. In contrast, we use business indicators for specifying what is to be achieved and how this achievement is to be done. The latter yields alternatives for which we discover relevant information to be put in the data warehouse. Thus it can be seen that the proposal of [Bar11] is for modeling/providing support to obtaining business intelligence whereas our interest is to use indicators for discovering data warehouse information contents.

Indicators have been used in requirements engineering [Lam09] for evaluating the contribution made by alternative sub-goals to the main goal. The idea behind this is to determine a qualitative degree of satisfaction of the top-level soft goals so as to select the option with the best degree. These indicators therefore define the choice of system functionality that is to be included in the system To-Be. In contrast, our use of indicators is for determining what targets an organization must set for its different work aspects. These indicators form the basis of determining choice sets which in turn help us to identify early information. There is no evaluation of any alternative. The purpose is to lay down a basis for discovering the decision making capacity to be supported and thereafter to reach the information to be maintained in the data warehouse.

We also looked at the area of decision making. However, we found that whereas our focus is on discovering information to be kept in the data warehouse, interest in decision making was on techniques to make the right selection from a choice set. Wang, Liu and Ruhe [Wan04] propose a cognitive process of decision making. The basis of their proposal is the trilogy of concepts, decision goal, a set of alternative choices, and a set of selection criteria or strategies. Al-Shawa and Basir [Sha09] talk about the special treatment needed for handling strategic goals as compared to tactical and operational goals. They focus attention on a reasoning framework about goals and plans to achieve strategic goals. Felix [Fel07] considers the problem of decision making in the presence of interacting goals. The proposal is to calculate interaction between goals using fuzzy types of interaction. Niu and Zhang [Niu08] propose a recognition primed decision model, RPD. The basic assumption of RPD is that when confronted with a decision situation, the decision maker will conduct an information processing process, which the authors call retrieval for decision. One kind of retrieval proposed by the authors is information defined as "what an individual needs to know for problem solving". The other is situation awareness. Niu and Zhang elaborate on the situation awareness aspect. As mentioned above, none of this work aims to define the information contents of a data warehouse.

6 Conclusion

Our proposal for early information requirements engineering starts off with targets, aspects and indicators. Indicators are the 'sticking glue' between the relevance and fulfillment dimensions, (i) they identify 'what is to be achieved' in the relevance hierarchy and (ii) can be used as a basis for defining 'how it is to be achieved' in the fulfillment dimension.

Once choice sets have been discovered in the fulfillment dimension, the issue is of determining the relevant information for alternative selection. This information is discovered through Ends, Means, Outcome Feedback, and Key Success Factor analysis techniques. The discovered information is 'early', abstract, fuzzy, and unstructured. It is to be converted to facts and dimensions in subsequent stages using the techniques spelled out in [Pra08].

The notion of target is relatively less fuzzy than that of goals in traditional goal orientation. This is because aspects and indicators are somewhat more concretely defined. This mitigates the first of the two problems of goal orientation pointed out by [Rol03]. Regarding target reduction, we have identified two mechanisms that can be followed. Thus, some guidance can be provided in the target reduction task. Again, this is to be contrasted with the difficulty of goal reduction in goal orientation.

Future work shall focus on elicitation techniques, elicitation of sub indicators and associated goals, choice set objectives and alternative ways of CSO achievement.

References

[Bar11] Barone, D., Jiang, L., Amyot, D., Mylopoulos, J.: Composite Indicators for Business Intelligence. In: Jeusfeld, M., Delcambre, L., Ling, T.-W. (eds.) ER 2011. LNCS, vol. 6998, pp. 448–458. Springer, Heidelberg (2011)

[Boe99] Boehnlein, M., Ulbrich vom Ende, A.: Deriving initial Data Warehouse Structures from the Conceptual Data Models of the Underlying Operational Information Systems. In: Proc. of Workshop on Data Warehousing and OLAP, DOPLAP, Kansas City, MO, USA (1999)

[Bon01] Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta, A., Paraboschi, S.: Designing data marts for data warehouses. ACMTOSEM 10(4), 452–483 (2001)

[Fel07] Felix, R.: Efficient Decision Making with Interactions Between Targets. In: Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Multicriteria Decision Making, MCDM 2007, pp. 221–226 (2007)

[Gol99] Golfarelli, M., Rizzi, S.: Designing the Data Warehouse: Key Steps and Crucial Issues. Journal of Computer Science and Information Management 2(3) (1999)

[Hor12] Horkoff, J., Yu, E.: Comparison and Evaluation of Goal-oriented Satisfaction Analysis Techniques. Requirement Engineering Journal (2012)

[Inm96] Inmon, W.H.: Building the Data Warehouse. John Wiley, New York

[Lam09] van Lamsweerde, A.: Reasoning About Alternative Requirements Options. In: Borgida, A.T., Chaudhri, V.K., Giorgini, P., Yu, E.S. (eds.) Conceptual Modeling: Foundations and Applications. LNCS, vol. 5600, pp. 380–397. Springer, Heidelberg (2009)

[Niu08] Niu, L., Zhang, G.: A Model of Cognition-Driven Decision Process for Business Intelligence. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 876–879 (2008)

[Pra04] Prakash, N., Singh, Y., Gosain, A.: Informational Scenarios for Data Warehouse Requirements Elicitation. In: Atzeni, P., Chu, W., Lu, H., Zhou, S., Ling, T.-W. (eds.) ER 2004. LNCS, vol. 3288, pp. 205–216. Springer, Heidelberg (2004)

[Pra08] Prakash, N., Gosain, A.: An Approach to Engineering the Requirements of Data Warehouses. Requirements Engineering Journal 13(1), 49–72 (2008)

[Pra09] Prakash, N., Prakash, D., Sharma, Y.K.: Towards Better Fitting Data Warehouse Systems. In: Persson, A., Stirna, J. (eds.) PoEM 2009. LNBIP, vol. 39, pp. 130–144. Springer, Heidelberg (2009)

[Rol03] Rolland, C.: Reasoning with Goals to Engineer Requirements, Invited talk. In: Int. Conf. on Enterprise Information Systems, ICEIS 2003 (2003)

[Sha09] Al-Shawa, M., Basir, O.: Constrained Rationality: Formal Targets-Reasoning Approach to Strategic Decision & Conflict Analysis. In: Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1439–1445 (2009)

[Wan04] Wang, Y., Liu, D., Ruhe, G.: Formal Description of the Cognitive Process of Decision Making. In: Proceedings of the Third IEEE International Conference on Cognitive Informatics, ICCI 2004, pp. 124–130 (2004)

[Yu97] Yu, E.: Towards modeling and reasoning support for early requirements engineering. In: Proceedings of the IEEE International Symposium on Requirement Engineering, pp. 226–235. IEEE Computer Society (1997)