Towards a Goal-Oriented Modeling Approach for Data Governance

Rohith Sothilingam [1], Vik Pant [1], Nazia Shahrin[1], and Eric Yu[1,2]

[1] Faculty of Information, University of Toronto, Toronto, Canada
[2] Department of Computer Science, University of Toronto, Toronto, Canada

Abstract
Data governance is a crucial capability in modern enterprises for quality decision-making and sound compliance management. Existing frameworks for data governance offer general guidelines and models but do not provide methods for responding to the specific requirements of an organization. This paper proposes goal-oriented conceptual modeling as a systematic approach for data governance that enables an organization to focus on their data-related objectives for generating insights and managing challenges while protecting their data. This approach builds upon existing modeling languages to complement generic data governance frameworks and achieve two outcomes. First, it provides a coherent way of connecting the mechanisms of data governance to enterprise governance objectives and business goals. Second, it makes data governance more understandable and communicable by enabling the exploration of alternatives, assessment of consequences, and uncovering of undesirable side-effects. The effectiveness of the proposed goal modeling approach is illustrated with a fictitious case study example.

Keywords
Conceptual Modeling, data governance, requirements engineering

1. Introduction

As organizations try to integrate and use large amounts of multifaceted data, having an effective data governance design becomes critical. A number of approaches for data governance have been proposed from different perspectives. Most existing frameworks consist of a set of principles together with guidelines for applying the principles when developing data governance solutions in organization. The IBM guideline for information Governance [4] and the DGI Data Governance Framework [11] are examples of frameworks which provide general prescriptive guidelines based on best practices. In [12], the authors offer guidance on how to organize a large enterprise into units where there are numerous data governance decisions to be made and who should be making such decisions. In [7], the authors provide a Big Data framework for handling urban governance issues and data analysis algorithms. The ISO/IEC 38500 framework [6] provides a set of general guideline with examples where one can identify various types of goals. However, the framework does not support the ability to conduct refinement or evaluation on such goals.

General guidelines often include checklists of best-practices which are intended to be applicable to any given organization. Checklists are an important element in measuring the level of adherence to data governance principles. Checklists help to map appropriate data governance procedures in an orderly, accountable manner.

Most existing approaches to data governance are prescriptive in nature and are meant to be applicable to any organizational setting. General guidelines require users to use their industry-specific knowledge, and personal experience and judgement to interpret the guidelines to suit their respective organization. Such approaches tend to assume that the prescription principles and guidelines can be applied straightforwardly, without consideration of competing goals and interests.

General guidelines do not offer a systematic way for each organization to determine how they can achieve their data governance goals within their business context. General guidelines do not provide support to determine whether a proposed mechanism will be enough, and if not, to explore how else to achieve the goals.

One approach to complement existing data governance frameworks is goal-oriented conceptual modeling. Goal modeling facilitates the traceability of decision rationale and can help manage change over time. For example, one may want to re-visit past decisions and understand the logic which constituted those decisions. A goal modeling approach follows a logical structure to visual representation of data governance goal structure. When making such decisions, conflicts and tensions between business and data governance objectives can surface, which require a resolution. For example, business objectives can often be focused toward revenue generation, whereas data governance objectives can often be aimed at preventing data from falling into the wrong hands. As a result, organizations commonly must balance the high cost of compliance with the importance of protecting enterprise data from potential compromises. Goal modeling helps to consider how data governance solutions will meet strategic business objectives, why the solutions are required, and how underlying tradeoffs among goals representing diverse strategic interests can be addressed through goal evaluation.

We propose a conceptual modeling approach using goals as the primary guide toward developing data governance schemes that would respond to the needs of the organization. We show how goal modeling can be used to explore mechanisms that enable data governance implementation in enterprise settings with simultaneous consideration of strategic business objectives. This paper aims to explore the ability for goal-oriented conceptual modeling to address the challenges of data governance, by answering two questions: *(1) "Can goal modeling be used to complement and provide added value to existing data governance frameworks?" and (2) "Can goal modeling provide the means to assess or evaluate alternative solutions and goals for data governance?"*

The remainder of this paper is structured as follows. Section 2 provides an example to motivate the research problem. Section 3 presents an overview of the benefits goal modeling as it relates to data governance. Section 4 provides an illustrated guide of goal modeling. Section 5 discusses limitations of proposed approach as well as insights identified. Section 6 summarizes related work and highlights the contribution of the proposed approach. The paper ends with concluding remarks and ongoing work directions in Section 7.

2. Motivating Example

Consider a fictional Financial Organization (FinCo). FinCo operates entirely in the cloud, specifically using Amazon Web Services (AWS). At FinCo, their financial services technology hosts customer data which must adhere to specific compliance standards followed by FinCo. Specifically, the organization's payment data must be compliant to the Payment Card Industry Data Security Standard (PCI-DSS).

There are 3 teams at FinCo that collaborate to achieve this objective: (i) the Data Governance team, (ii) the Data Steward Team, and (iii) the Business Project team. Each team has their own strategic interests, along with their shared interest to achieve FinCo's strategic business objectives.

The Business Project team is responsible for responsible for helping FinCo achieve specific results which helps the business maintain and grow. This team aims to maximize revenue and improve customer service excellence, while maintaining a cost management discipline. Such goals are examples of *Business Strategy* goals.

The Data Governance team is responsible for the ownership of the payment data, compliance of the payment data, and are responsible for adhering to data policies for access, storage, and administration. This team aims to ensure all data that is stored and supporting data procedures are PCI-DSS compliant. Such goals are examples of *Compliance* goals. The Data Governance team looks to achieve the satisfaction of compliance goals through the enforcement of policies. Adherence to policies is monitored through regularly scheduled internal reviews to ensure encryption policies are up-to-date according to the PCI-DSS standard. Such goals are examples of *Policy* goals.

The Data Steward team is responsible for collecting, organizing, and managing challenges and problems with data. The Data Steward team aims to achieve well-managed, secure data that is compliant to PCI-DSS, and is efficient for data analytics. The team looks to accomplish this with a resilient database management system (DBMS) for short and long-term storage. Such goals are examples of *Data Strategy* goals.

As such goals are defined, potential conflicts may arise due to having various options for achieving goals. Such conflicts lead to the need for making tradeoffs between conflicting goals. It is important to consider the tradeoff between conflicting goals to reach the best approach for a solution. As the Data Governance team focuses their attention on compliance of underlying data, the Data Steward team looks to emphasize the importance of highly efficient database management and data quality. Both goals are highly costly and can ultimately bring out reluctance from data stewards to compromise efforts and costs for database management for the sake of achieving compliance.

To ensure this conflict can reach a fruitful resolution for FinCo, the Data Governance team, Business Project team, and Data Steward team must work together. To assess how to resolve the conflict, the teams explore alternatives and analyze the tradeoffs , with the help of goal modeling.

The Data Governance team can choose to conduct formal internal reviews to emphasize the adherence to PCI-DSS compliance. This solution will impose additional processes on the project team which would increase overall costs as well as the effort and time spent toward database management. The following question is faced: "How can the project team compare the net benefit from this option rationally?"

As an alternative, the Data Governance team suggests FinCo can invest in automated compliance services in the cloud which are available to them. Specifically, the team suggests AWS Config, which is a service that monitors and alerts for compliance conformance throughout FinCo's cloud ecosystem, in accordance to the latest PCI-DSS standard. This alternative is faced with reluctance from the Business Project and Data Stewardship teams due to the high costs involved with this solution. The following question is faced: "How can the team trace decision rationale to systematically compare these tradeoffs?".

By collaboratively analyzing the conflicting goals and potential tradeoffs, the teams discover that the benefit of investing in AWS Config as a more vigilant and accountable compliance solution outweighs the higher cost for FinCo. This motivating example, though fictitious, shows how goal modeling can potentially help the FinCo to express and analyze their tradeoff of goals and options available for solutions. Taking a goal modeling approach can help real-world organizations to evaluate alternative data governance solutions by assessing tradeoffs which emerge from other types of goals which conflict with data governance goals, to make better decisions. .

3. Potential Benefits of Goal Modeling for Data Governance

This section outlines the benefits and areas of guidance that goal modeling can provide to achieve data governance in a systematic way.

Contextualized guidance for data governance. Using goal refinement, goal modeling achieves contextualized guidance by helping to clarify governance principles and guidelines, and to interpret them according to the specific needs and circumstances of the organization. Through goal propagation, goal modeling offers a systematic way of achieving and evaluating the data governance goals of an organization. Goal propagation provides the ability to assess whether a proposed mechanism will be enough. A goal-based model facilitates the ability to connect data governance processes and mechanisms to enterprise goals and requirements to understand the sufficiency of such processes and mechanisms. The method can help governance teams and business leaders arrive at common understandings and to seek collaborative solutions.

Traceability and decision rationale. The main goal modeling structures demonstrated in this work consist of goal refinement and the operationalization of those goals to provide actionable solution mechanisms. Goal modeling helps with principled and disciplined decision-making by systematically linking data assets and business value. By linking strategic goals to more granular decomposed goals through refinement, organizations can better trace the rationale which constituted how solutions were

identified or compared. Providing a traceability of decision rationales helps to manage changing requirements over time. For example, organizations can re-visit past decisions and understand what criteria constituted those decisions, through refinement and goal evaluation.

Communication and collaboration within and across teams. In addressing data governance, requirements can emerge from various teams with conflicting strategic interests such cost, compliance, and security. As such, tradeoffs emerge, and conflicts must be resolved. Goal modeling can identify tradeoffs by uncovering and providing the means to evaluate the suitability of conflicting solution mechanisms. As a result, goal modeling can resolve conflicts which emerge from such tradeoffs. Goal modeling can help teams communicate and facilitate negotiation to achieve the most agreeable outcome. Goal modeling can improve accountability of the decisions, if agreed to through explicit shared discussion among relevant parties. In a large organization, the explicit reasoning can serve to justify decisions to higher authorities, to coordinate with peer units, and to guide downstream implementation. Such reasoning can advance the learning of the organization, so that similar Data Governance decisions in the future can benefit from the documented experience.

4. Related Work

Goal modeling structures are well known in business modeling (e.g., [9]) and have been widely investigated (e.g., [5]). However, to the best of our knowledge, the application of systematic goal-oriented methods and conceptual modeling is novel within the area of data governance.

Recently several frameworks addressing data and information governance have been developed highlighting goals or conceptual modeling. In this section, we acknowledge the contributions of past approaches and discuss the differentiating benefits of the proposed framework.

In [10], the authors present a methodology for Business Information Modeling that can be used to analyze data governance requirements of organizations. They offer a glossary of modeling concepts related to data governance that are expressed within Entity Relationship diagrams. This approach provides a focused mapping of data governance concepts to support the analysis of data governance requirements. While this methodology includes a collection of fixed goals, it does not offer a systematic means for adding new goals or adapting the meanings of fixed goals to fit user requirements. Similarly, in [1] the authors define six focus areas as data decision domains: data quality, data security, data architecture, data lifecycle, meta data, and data storage & infrastructure. Narrowing specific focus areas allows the authors to achieve sufficient coverage of specific areas of data governance. Though the authors do not claim to provide an exhaustive approach, the challenge with this approach is that it does not account for domains outside of the chosen focus areas. Given the benefits of approaches which address specific domain within data governance, our proposed approach offers a different perspective, instead emphasizing the ability to support the inclusion of as many types of data governance objectives and domain areas as needed.

In [8], the authors propose a taxonomy of data governance objectives within ecosystems and offer a catalog of goals to depict various characteristics of data governance in a tabular format. We acknowledge the benefits of this work and despite the benefits gained from the catalog, this approach is limited to the expression and analysis of model content in a table rather than in a visual diagram which explicitly connects data governance and business objectives. As a perspective which provides visual guidance, the proposed framework offers a modeling approach that results in visual diagrams that complement existing data governance frameworks and can represent a richer set of content in a more intuitive and understandable manner.

5. Goal Modeling Process

In this section, we outline a step-by-step guide to the proposed approach, which comprises two phases: Modeling and Evaluation & Exploration. To demonstrate each step, we describe examples from the FinCo example.

5.1. Modeling Steps

The modeler develops a goal model using the concepts defined in the Data Governance framework adopted by the organization. These concepts are interpreted as specialized types of goals or operationalizations in constructing the goal model. Data Governance frameworks may vary in their terminologies and concepts and may or may not have a metamodel defined. In the example, we assume that the framework includes the following: *business strategy*, *data strategy*, *policy*, and *compliance plan*. In the figures, these are marked with annotations 'B', 'D', 'P', and 'C'. These informal annotations guide the model elaboration, but have no effect on goal model evaluation. During each step of this phase, goals are decomposed until sufficiently granular enough to link to the next specialization of goal or operationalization.

M1. First, identify strategic enterprise objectives which the organization is aiming to achieve. Continue to decompose such objectives into further underlying objectives to bring out details beyond the original objective. These goals will be considered *Business Strategy* goals. FinCo's key project success criteria are the following: (1) sustainable business, (2) improving customer service excellence, and (3) maximizing revenue. These success criteria are represented as *Business Strategy* goals.

M2. Next, identify requirements for adherence to compliance standard(s) and the plans for how such standards will be adhered to. Examples of compliance standards include PCI-DSS for financial institutions and examples of plans for governance include privacy legislation. Compliance measures will be considered *Compliance Plan* goals. The following *Compliance Plan* goals can be put in place to support *Business Strategy* goals and adherence to PCI-DSS: "Customer data is PCI-DSS compliant". This goal is then decomposed into further *Compliance Plan* goals.

M3. Next, identify strategic objectives on how underlying data will be managed, of which are to be instrumental to achieving objectives which manage enterprise data. These goals will be considered *Data Strategy* goals. The following *Data Strategy* goals can be put in place as a strategic measure to contribute to its success of *Compliance Plan* goals: "Database management be vigilant" and "DBMS is resilient and encrypted". The goals are then decomposed into further *Data Strategy* goals.

M4. Next, identify strategic objectives on policies which can be instrumental to enforcing compliance and governance objectives. These goals will be considered *Policy* goals. To address the previously mentioned *Compliance Plan* Goal, the following *Policy* goal can be put in place in addition to the related *Data Strategy* goal to ensure a well-defined goal is specified for achieving PCI-DSS compliance requirements: "Encryption policies be up-to-date". This goal is then further decomposed into further *Policy* goals.

M5. Finally, identify solution mechanisms for goals which have been sufficiently decomposed, which are to be considered "leaf" goals, and attach the solution mechanisms. These solution mechanisms are considered as *Operationalizations*. The specialization of *Operationalizations* must be the same as the respective goal it is attached to. Alternative solutions may be present, which can be conflicting and cause tradeoffs.

In the case of FinCo, two instances of tradeoffs exist (Fig. 1. In the first case, FinCo can achieve PCI-DSS compliance by mandating formal compliance reviews. In the second case, FinCo can invest in the costly AWS Config service. The modeling is complete with the first pass of the Modeling Steps once all intended goals and Operationalizations have been sufficiently decomposed and expressed. The modeler can re-visit the Modeling Steps iteratively if any changes are needed.
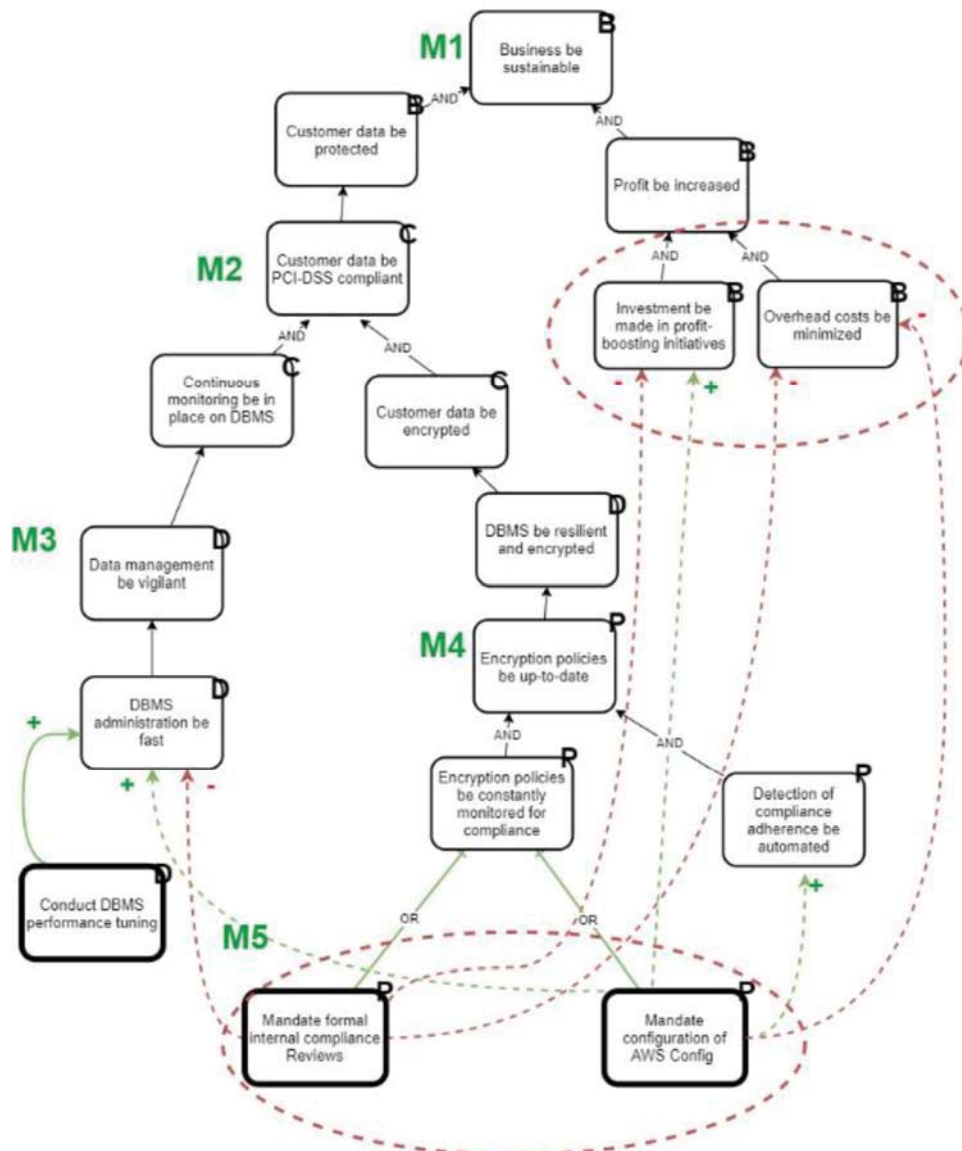
Fig. 1 Identifying solution mechanisms and tradeoffs

5.2. Evaluation & Exploration Steps

In this series of steps, the modeler analyzes the instantiated goal model that is developed in the Modeling steps. The modeler evaluates the satisfaction of goals by tracing the contributions of operationalizations and using propagation labels. The modeler may re-visit the modeling to reconsider goals until satisfied. The modeler evaluates goal satisfaction of the goals created in the Modeling Phase by propagating labels. Goals can either be: fully satisfied (denoted by a checkmark), partially satisfied (denoted by an "X" with a dot underneath), fully denied (denoted by a cross), or partially denied (denoted by a dot underneath an "X"). Satisfaction of goals is determined by the satisfaction of *Operationalizations* and the chain of satisfaction from subsequent decomposed goals until goal satisfaction labels are applied to the highest-level goal (i.e. *Business Strategy* goals). In case the status of a softgoal or goal cannot be resolved then it can be marked as unknown (denoted by an "X" by a dot with a question mark). This goal modeling notation and evaluation scheme is adopted from the Goal-oriented Requirements Language (GRL) modeling language [3].

If alternatives exist, conduct propagation on each alternative to compare the benefits of each. As shown in Fig. 1, two alternative operationalizations are available to achieve the *Policy* goal of "Encryption policies be constantly monitored for compliance". In Fig. 2.1 and Fig. 2.2, we conduct goal propagations on each alternative. In the case of Alternative 1 (Fig. 2.1), the top strategic objective is

not met. In Alternative 2 (Fig. 2.2), the top *Business Strategy* goal of "Sustainable Business" is fully satisfied at the cost of "overhead costs be optimized" not being satisfied. In the next step, the modeler must explore further alternatives by considering how else some of the goals in the model can be achieved.
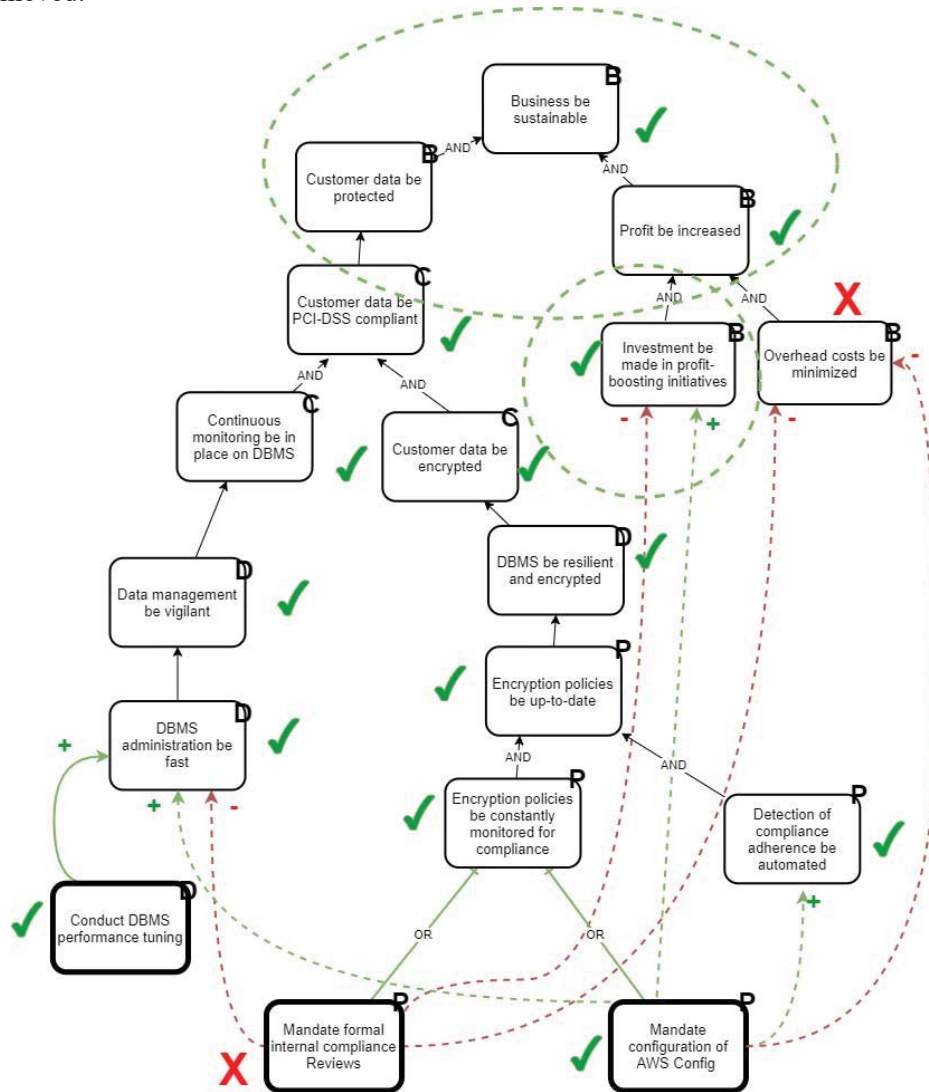

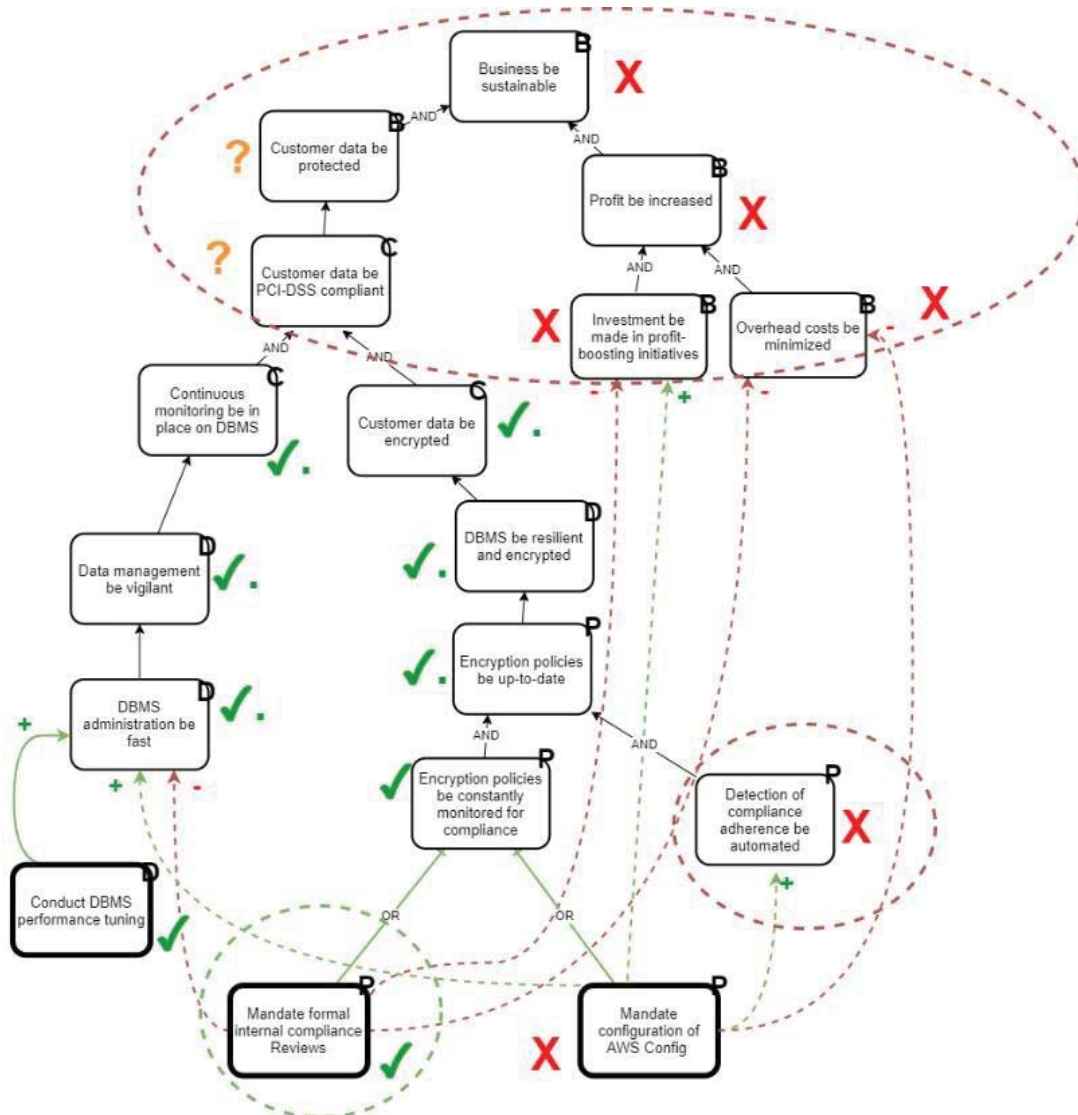**Fig. 2.1 Alternative 1: Cloud-based service for compliance monitoring**

Fig. 2.2 Alternative 2: Mandated formal reviews for compliance monitoring

6. Discussion and Limitations

In this section, we identify several important insights from the modeling and discuss the usefulness of the approach for data governance.

Looking at evidence of the usefulness of recursive goal refinement, Fig. 1 demonstrates an example of a scenario where goal modeling can trace the linkage between chosen solutions (i.e. *Operationalizations*) to high-level business objectives (i.e. *Business Strategy goals*). This was shown through the decomposition of subsequent underlying goals until a solution is sufficiently granular enough. As a result, this example demonstrated that in such a scenario, goal decomposition can add value to existing Data Governance frameworks by improving traceability of decision rationale. Like the example of FinCo, goal decomposition can be used to help tailor Data Governance frameworks to a given organization, as guided through goal-based evaluation. The motivating example demonstrated that goal modeling can be used to better understand what options were considered and what underlying network of goals were affected. As a result, an organization can revisit the decision rationale to consider other means of data governance objectives.

The comparison of alternatives demonstrated in Section 4 shows the detailed rationale of *how* the investment in AWS Config service would yield greater direct benefit than the alternative solution of formal compliance reviews. It can be seen in Fig. 2.1 and Fig. 2.2 that in this scenario, goal modeling can be used to evaluate alternative data governance solutions and goals, by exploring alternative

Operationalizations. Specifically, the modeler can see what data governance goals, as related to specific business objectives, constituted the rationale of the decision. In similar organizational scenarios, this method can potentially help such governance teams and business leaders arrive at common understandings and to seek collaborative solutions.

Given the benefits of the proposed approach, we acknowledge that the goal modeling demonstration is based on a contrived example. To solidify the constituent features of the proposed approach, empirical research in a real-world setting to better test the approach. In addition, an explicit and formal metamodel has not yet been proposed as part of the goal modeling approach.

7. Conclusions and Ongoing Work

This paper demonstrated an example of how goal-oriented conceptual modeling can be helpful as a systematic approach for implementing enterprise data governance. In future work, we plan to fully define a metamodel and elaborate on the modeling steps. We plan to investigate how to use goal modeling to complement existing Data Governance frameworks, many of which do not have explicit metamodels for their information structures. Having a metamodel would enable us to develop tool support for model-based Data Governance. We plan to incorporate quantitative evaluation capabilities using quantitative indicators. For example, in [2] the author uses key performance indicators (KPI) with a GRL profile to measure compliance levels. We will consider including justifications for alternatives, using for example the "belief" concept in i* [13]. We plan to engage organizations to test the goal modeling approach in real-world settings and uncover what obstacles may be encountered.

8. References

[1] Abraham, R., Schneider, J., & Vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*, 424-438.

[2] Akhigbe, O. (2016). Towards a Regulator-Oriented Regulatory Intelligence Framework. In *2016 IEEE 24th International Requirements Engineering Conference (RE)* (pp. 415-420). IEEE.

[3] Amyot, D., Ghanavati, S., Horkoff, J., Mussbacher, G., Peyton, L., & Yu, E. (2010). Evaluating goal models within the goal-oriented requirement language. *International Journal of Intelligent Systems*, *25*(8), 841-877.

[4] Ballard, C., Compert, C., Jesionowski, T., Milman, I., Plants, B., Rosen, B., & Smith, H. (2014). *Information governance principles and practices for a big data landscape*. IBM Redbooks.

[5] Horkoff, J., Barone, D., Jiang, L., Yu, E., Amyot, D., Borgida, A., & Mylopoulos, J. (2014). Strategic business modeling: representation and reasoning. *Software & Systems Modeling*, *13*(3), 1015-1041.

[6] ISO/IEC TR 38505-2:2018(2018) Information technology — Governance of IT — Governance of data — Part 2: Implications of ISO/IEC 38505-1 for data management. (2018). Retrieved February 24, 2021, from https://www.iso.org/obp/ui/#iso:std:iso-iec:tr:38505:-2:ed-1:v1:en.

[7] Ju, J., Liu, L. and Feng, Y., (2018), Citizen-centered big data analysis-driven governance intelligence framework for smart cities, Telecommunications Policy.

[8] Lis, D., & Otto, B. (2021). Towards a Taxonomy of Ecosystem Data Governance. In Proceedings of the 54th Hawaii International Conference on System Sciences (p. 6067).

[9] Object Management Group: Business Motivation Model, Version 1.0. Technical Report May, Object Management Group (2002). http://www.omg.org/spec/BMM/1.0/.

[10] Priebe, T., & Markus, S. (2015). Business information modeling: A methodology for data-intensive projects, data science and big data governance. In 2015 IEEE International Conference on Big Data (Big Data) (pp. 2056- 2065). IEEE.

[11] Thomas, G. (2006). The DGI Data Governance Framework: The Data Governance Institute. *Orlando, FL, USA*.

[12] Weber K., Otto B., Österle H. One size does not fit all—a contingency approach to data governance Journal of Data and Information Quality (JDIQ), 1 (1) (2009), p. 4.

[13] Yu, E., Giorgini, P., Maiden, N., & Mylopoulos, J. (Eds.). (2011). Social modeling for requirements engineering. MIT press.