Lean Six Sigma Roadmap for Quality Assurance of Biomedical Ontologies

Gabriele Arcidiacono
Department of Innovation and Information Engineering (DIIE)
Guglielmo Marconi University, Via Plinio 44, 00193 Rome, Italy.

ORCID iD: 0000-0002-7712-1009 & Scopus Author ID: 56656284600

Isabella Martini1 and Ernesto. W. De Luca2
1Leanprove, via La Marmora, 45, 50121 Florence, Italy.
2Department of Innovation and Information Engineering (DIIE)
Guglielmo Marconi University, Via Plinio 44, 00193 Rome, Italy.

Abstract
The purpose of this article is to introduce the Lean Six Sigma (LSS) DMAIC (Define, Measure, Analyze, Improve, Control) roadmap for quality assurance in biomedical ontologies, by applying Lean Six Sigma principles to Ontology Engineering and Collaboration Engineering. Collaboration lies at the human core of social processes and interactions, where a shared goal is what participants join efforts to achieve; therefore, combining Collaboration Engineering and LSS, and, specifically, the DMAIC route which structures each LSS project aims at creating an applicable standard for quality assurance. That will help establish effective biomedical interfunctional databases that can help to improve health care processes, answering the need to reduce the amount of human errors associated with inefficient health care processes.

The research idea suggested within this paper aims at developing ontologies that are functional to meet health care needs, by ensuring data collection reliability, by creating a common language to speak with data, and by promoting a shared culture of process optimization and cost effectiveness. Semantic approaches for Collaboration Engineering have been debated in previous works, where a new ontology-based approach has been proposed, where collaboration knowledge is collected, managed and shared by connecting each concept of an ontology to a more detailed collaboration step or a resource, by applying LSS principles.

The originality of this article lies in introducing LSS DMAIC as an effective and efficient scientific methodology to query and feed an ontology designed for being used in health care, and, at the same time, as a tool for quality assurance of biomedical ontologies as well.

INTRODUCTION
With the onset of Industry 4.0 and its related assets, such as the concept of smart factory, the Internet of Things, and Big Data, increasingly efficient shared processes of automation and digitization have revolutionized the use of data and information technologies, ultimately aiming at connecting, innovating and manage supply and value chains. Such framework has met the need for the biomedical sector to be ahead in the continuous innovation and improvement of its organizational system processes and procedures. In particular, digitizing biomedical processes has improved the management of pathologies, with more patient-adapted treatments and monitoring. Moreover, biomedical research has taken advantage of patient data digitization, in order to develop medicines tailored to patient needs, and to achieve higher standards of efficiency in treating conditions.

Therefore, when dealing with health care processes, it has become increasingly evident that human error needs to be reduced. Various solutions have been attempted and experimented so far, with a variety of different approaches, such as among the many others, the recent one suggested by Collinge [1] and debating the integration of social semiotics and design work practice. Since one of the most crucial needs is to ensure efficient communication within hospitals to have the system work, one solution to be explored is to ensure the entire medical staff (doctors, nurses, laboratory technicians) the availability of standardized shared tools to speak a common language. For its potential to investigate processes and phenomena in their smallest details, Lean Six Sigma (LSS) is debated here as a methodology to standardize the procedure to create an ontology, and to keep contributing to build it on a regular basis. The novelty of this article is the use of LSS DMAIC (Define, Measure, Analyze, Improve, Control) roadmap [2] as scientific methodology to query and feed an ontology designed for being used in health care, and, at the same time, as a tool for quality assurance of biomedical ontologies as well.
The article is divided in five sections. Section two introduces a comprehensive literature review on ontologies as instruments for knowledge representation, on research methodology for Ontology Engineering and for Collaboration Engineering. Section three introduces the proposed methodology, i.e. DMAIC explained as tool for quality assurance of biomedical ontologies. Section four relates on some successful applications of DMAIC to deal with health care process optimization, with a focus on criticalities in knowledge and information exchange. Eventually, section five offers some conclusive remarks and research perspectives on the application of DMAIC, and LSS, for quality assurance of biomedical ontologies.

**LITERATURE REVIEW AND RESEARCH SCOPE**

A. Capturing Semantics in Ontologies

Ontologies are an important instrument for knowledge representation. They have been studied and applied in the context of many fields such as within Knowledge Engineering, Digital Libraries, Artificial Intelligence, or Information Science. Ontologies have been defined by Gruber [3] and are still considered fundamental in the field of Ontology design and Engineering, and for knowledge sharing. Different work on ontology learning addresses specific methods according to the type of information sources included to extract knowledge [4]. Knowledge Engineering methods focus mostly on description logics and formal foundations of ontology design [5].

In this article, human-centred and collaborative activities play an important role in the cognitive design of ontology facets. The proposed approach is likely to contribute to the further analysis of human-centred ontology design methodologies.

The concept of ontology here describes semantics that can be used for different purposes, such as for Information Retrieval. This representation is mostly based on texts which is well understood and describe semantic relations between concepts. The research in this area comprises mainly semantic search models [6] and user studies [7], however, mainly in technical and natural scientific domains, whereas research addresses the humanities with considerably less attention.

Furthermore, recent research activities apply fuzzy concepts in order to reduce the so-called semantic gap [8] [9], described by Hein, Goldberg, Michelfelder [10] as “the difference in meaning between constructs formed within different representation systems”. On the other side, Zarka, Ammar, and Alimi [11] propose a framework and workflow which describes a two steps approach constructing a fuzzy ontology and analyzing learning dataset. This has been used for an annotation process through a reasoning engine.

Analysing the health and medical care data, we recognize that interoperability issues are seen as the major challenge in this research field. The data coming from different sources is stored in different and distributed data repositories. In addition, there are different administrative domains which are inconsistent in naming, structure and/or format. Therefore, it is difficult to make data easily accessible, accurate and manageable in terms of efficient data processing and integration with other systems [12].

Also in this research field ontologies can play an important role. Several health care domain ontologies have been already created describing specific domains in biomedicine (e.g. within parameters like anatomical parts and their relations, or terms used in clinical medicine, or rehabilitation domain). They are comparable for some properties and most of them are available via BioPortal [13]. An overview is given in Table 1.

<table>
<thead>
<tr>
<th>Name of the ontology</th>
<th>Number of classes</th>
<th>Number of properties</th>
<th>Number of projects</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNOMEDCT</td>
<td>303035</td>
<td>152</td>
<td>21</td>
<td>UMLS</td>
</tr>
<tr>
<td>LOINCS</td>
<td>173271</td>
<td>111</td>
<td>5</td>
<td>UMLS</td>
</tr>
<tr>
<td>MedDRA</td>
<td>65934</td>
<td>16</td>
<td>4</td>
<td>UMLS</td>
</tr>
<tr>
<td>CTCAE</td>
<td>2000000</td>
<td>N/A</td>
<td>1</td>
<td>OWL</td>
</tr>
<tr>
<td>FMA</td>
<td>100080</td>
<td>188</td>
<td>15</td>
<td>OWL</td>
</tr>
<tr>
<td>ICD10</td>
<td>12451</td>
<td>1</td>
<td>3</td>
<td>UMLS</td>
</tr>
<tr>
<td>RADLEX</td>
<td>46059</td>
<td>95</td>
<td>8</td>
<td>OWL</td>
</tr>
</tbody>
</table>

In order to overcome the interoperability problem ontology alignment and matching techniques, as recently reviewed by Otero-Cerdeira, Rodriguez-Martínez, and Gómez-Rodríguez [14] can be applied in the proposed work, since multiple ontologies can be expected to be of relevance to the description of entities in health reports as well as for the representation of appropriate ontology facets.

Furthermore, as Khodaskar and Ladhake [15] discussed, ontology alignment promises to improve image retrieval by means of more than one integrated ontology.

An image can be expressed in text by annotating it in relation to the object around it. At the moment, research in this field focuses on text-based image retrieval, content-based image retrieval, and ontology-based image retrieval techniques [16] [17] [18]. Lew, Sebe, Djeraiba, and Jain [19] name the major challenges for the content-based method, notably inconsistency as the major issue in using folksonomies. As another approach, the research area of Semantic Image Interpretation looks at the ontological description of images. Two prominent approaches follow logic-based techniques and recently also neural networks-based approaches, including promising deep-learning methods [20]. Thus, the use of such semantically annotated images can be helpful for doctors and nurses, just to mention an example of its applicability.
B. Research Methodology for Ontology Engineering

Ontology Engineering aims at building a formal representation of domain knowledge (concepts in a domain) and creating a common understanding of the structure of information in the domain (relations between the concepts) among people or software agents [21] [22] [23]). As already presented above, today, several methods and methodologies for developing ontologies already exist [24]. Uschold and Gruninger [25] presented a methodology for Ontology Engineering. These methodologies were adopted for ontology building [26] [27]) and for Collaboration Engineering process [28]. For this purpose, this approach was used to develop our Collaboration Engineering ontology and will adapt it for supporting doctors and nurses in their data-driven analysis. The purpose of the collaboration ontology is to describe collaboration from an external point of view and adapt it to the different views (in our case the different facets being dependent, for example, on the expertise).

Sure, Staab, and Studer [29] show some more specific concerns of Ontology Engineering related in the overall process. Different processes have to be considered implementing and launching a Knowledge Management Application, as shown in Figure 1.

![Figure 1: Relevant processes for developing and deploying Knowledge Management Application](image)

Three major processes have been considered in this work: “Knowledge Meta Process”, “Human Issues” and “Software Engineering”. They have to be considered interdisciplinary and help to solve problems by automated IT solutions and, at the same time, by human actors and through organizational processes.

The Knowledge Meta Process can be considered to be the core process of Ontology Engineering. Software Engineering (SE) helps in accompanying the other processes for knowledge management applications. While “Human Issues” could dominate the processes depending from the domain knowledge of the experts. Here, collaborative Ontology Engineering requires physical presence and advanced tool support to be very helpful for early stages of Ontology Engineering [29] as discussed in the following.

C. Collaboration Engineering

On the other hand, collaboration is very important in many aspects of our lives. When we suppose somebody works together in an hospital, they want to reach goals faster, yield some joint results and inspire each other during their collaboration activities. The synergy effects can boost all kinds of achievements tremendously. Anyhow, some collaboration efforts could not work. Thus, the assistance, analysis and support of collaboration processes with technological means could be very innovative. With this work, we aim to explore this methodology in a real world use case and give best practices on how ontologies could be used for the given collaboration process in health care.

We have analysed semantic approaches for Collaboration Engineering in a previous work [27] presenting a new ontology-based approach, where each concept of the ontology is connected to a more detailed collaboration step or a resource, to collect, manage and share collaboration knowledge. In this work we presented the utility of the proposed ontology in the context of a real-world scenario where we describe the model of a collaboration process and how this can be supported by using our ontology. In addition, we discussed how well-known ontologies, such as the Friend Of A Friend (FOAF) ontology [30], can be linked to our ontology and extend it. While the focus of the work was on semantic Collaboration Engineering, we additionally presented methods of reasoning and data mining to derive new knowledge on the collaboration process as a further research step.

Adapting the methodologies for the health care domain, we aim to support people in decision making to reduce the possibility to avoid common mistakes and to help in a data-driven way the analysis of a given case.

By using an ontology, we can increase processes information about collaboration. Information about the patient characteristics, like the symptoms of the persons can be used to define formalized rules for group composition of patients. Context information further provides the possibility to adopt the design pattern approach also for the whole collaboration process. Thereby the objective and the description of the client can be used to define selection guides for a generic collaboration process. This modelling decisions have to be taken by the Collaboration Engineer depending on the Web Ontology Language (OWL)-Format [31] variant to be used for the reasoning tasks and for providing access and support to the nurses and doctors. Hereby visualisation processes and text-based recommendations could be helpful for this kind of work.
METHODOLOGY: DMAIC FOR QUALITY ASSURANCE

This section introduces the research methodology singled out as roadmap to create a standard for semantic data acquisition functional to health care contexts. Collaboration Engineering is here combined with LSS – specifically, the DMAIC route which structures each LSS project [32] – as shown in Figure 2. This aims at creating an applicable standard for quality assurance, to help establish effective biomedical interfunctional databases. Such newly defined standard will improve health care processes, answering the need to reduce the amount of human errors associated with inefficient health care processes. This because Lean, and therefore LSS, is suitable to be adapted to any process with a certain degree of complexity, where the human component is affected by the process. Customer value is the driver of Lean and of LSS in several sectors. Starting from manufacturing [33], to small and medium enterprises [34], LSS is successfully applied in in any sector, as recently demonstrated also by Nicoletti [35], who discussed LSS and digitize methodology applied to management of data and to communication flow, in what has been named “Lean procurement”.

Moreover, the Multi-Level Hierarchical (MLH) modelling technique [39] has been used to determine the dependence of a system reliability [40], and can be extended to the design matrix of the Project-based Learning (PBL) model.

What makes LSS intervention particularly effective is indeed its PBL approach, because it focuses the learning processes on structured problem solving. Instructors then become process facilitators, and guide learners while they explore the challenges they will encounter when applying learning subjects to real-life work (but not only) situations. PBL starts from five questions, formulated by the participants of the learning process, and articulated in a most complete series of hypotheses and explanations [41]. DMAIC usually deploys along five steps too, which structure a LSS project for process optimization aimed at Process Excellence. DMAIC provides the logic underneath any LSS optimization process, and can be used to create a standard for quality assurance when building an ontology. This because ontology data and information need to be quickly accessible, to provide proper answers, avoiding information gaps, and, above all, to be collected meeting well-identified criteria. DMAIC serves well this purpose, as it allows to create increasingly complex database query systems. By applying DMAIC, organizational technical memories are created, whereby, when a problem is mapped and solved, the solution is recorded and stored to be accessible by all members of the organization, and an internal information repository is created that keeps increasing with new data, gathered with a systematic and standardized approach. Internal, and shared, information repositories thus created are also a collection of experiences dealing with similar problems.

Therefore, DMAIC ideally becomes the framework to create the logic lying at the core of data and information acquisition, collection, and process for any ontology. This because DMAIC questions the system for the problems (Define: which problems are there?), speaks with data (Measure: what data do we collect?), searches and retrieves the root causes of problems (Analyze), heads towards a solution of the problem (Improve) and establishes procedures to monitor that the critical issues no longer occur (Control). DMAIC compares and contrasts data statistically, clustering and reading them in the most objective way, paying attention to single data, and not only to the generic symptoms of a problem. This would ensure a set of measurable standards for quality assurance in ontology creation. Most importantly, DMAIC could set the creation process so as to ensure that ontologies speak a common language, that is understood by different professional categories: health care professionals would specifically benefit from a quality assurance standard for its shared semantic architecture, which will ensure it to comply with established criteria of accessibility and readability for a wide range of users.

To understand how DMAIC could be applied for quality assurance to biomedical ontology, it is worth providing more details on each phase of the cycle.

The Define phase of the DMAIC cycle plans the Six Sigma
intervention, and defines its framework of reference, whereby it must specify the problem to be analysed; the potential process clients and their needs; the Key Process Indexes (CTQ, Critical to Quality); the goals to be attained with the LSS project; and the resources needed and time to conclusion (Project constraints).

Moreover, process mapping allows to describe in details the process under scrutiny, to spotlight its criticalities, and its value-added (VA) and non-value-added (NVA) activities. Through a visual representation it is possible to provide a univocal and shared description of the process. This allows to understand how processes are deployed within each organisation, determining the role of each participant.

Measures is the second step of a LSS project, where a “rational” Data Collection is performed for the chosen scope of the intervention: this collection requires effective and efficient planning in order to create a database of knowledge to record the process which will highlight the critical issues from an objective standpoint. Data are interpreted through statistical tools, and the reliability of data is verified. Process Performance is calculated through the proper KPI (Process Sigma, Process Capability etc.). Typical activities are collecting the necessary data to fully understand the problem; focusing attention on improving the collection of data relevant to the present situation; implementing the measurement systems; analyzing the process capability; and using descriptive statistics.

Creating a structured data collection plan is crucial to detect the information that are needed for performance analysis. Data analysis helps to tell the difference between what we think is happening and what is actually happening; to confirm or refute ideas, biases, and theories; to establish a ground level for performance; to identify and understand relations that could explain variability.

The Analyze phase aims at identifying the root causes of the problem and at quantifying (although sometimes approximately) their influence on CTQs. Therefore, it is necessary to identify the causes that contribute to a variation of the process output (defect/problem); to identify the cause-effect relations; to select the main causes that influence the result, to so-called root causes. Six standard categories can be applied to arrange and manage potential root causes, i.e. man, method, machine, measurement, material, mother nature – which include all the causes which are out of human control.

The Improve phase is centred on the search for solutions to improve the problems detected in the previous phases, in order to optimize the process under scrutiny. Therefore, it is necessary to develop improvement opportunities, by developing, trying and taking solutions which remedy the causes found in the previous phase; to choose the most influential factors; to develop a “pilot test”; and to define how to evaluate results in the Control phase.

The Control phase evaluates the effectiveness of implemented solutions and assures the sustainability of the improvement over time, thus standardizing the process. Activities usually featured within this phase are standardizing and documenting effective methodologies; monitoring process performance over time (Statistical Process Control); evaluating the results/benefits of the process; and sharing acquired knowledge.

RESULTS: SUCCESSFUL APPLICATIONS OF DMAIC IN HEALTH CARE

Out of experience in health care context of implemented LSS projects in the past years in Italy, but not only, the main criticalities appeared to occur on the lack of a shared system to communicate within hospitals and health care facilities and institutions [42]. This had led to issues that the authors were asked to correct with their intervention, but, at the same time, stimulated further research threads to connect statistics engineering and Collaboration Engineering to health care and, ultimately, social improvement.

One intervention was specifically aimed at optimizing costs for patients suffering from chronic diabetes. Among the objectives of the LSS projects were to harmonize procedures, clinical pathways, and shared language among GPs and hospital consultants; to identify clinical pathways for «out-patients»; to reduce unneeded hospitalization and to decrease hospitalization mean time, thus positively impacting on costs and waiting lists. All this aiming at providing a better care for patients affected by chronic diseases.

As singled out of the Measure and Analyze steps of the DMAIC cycle, the main criticalities were indeed related to a lack of a shared system to communicate real-time data on the same patients to all the departments involved in their treatment. Namely, the criticalities were:

- the lack of a standard protocol managing the various types of clinical operations;
- lack of appropriate communication among GPs and hospital consultants, leading to inappropriate prescriptions;
- low process standardization;
- long waiting times between booking a specialist medical examination and the actual examination.

A second LSS intervention was instead focused on improving the quality of hospitalization scheduling and management for patients affected by rheumatic conditions. Specifically, the clinical pathway leading to hospitalization for chronic rheumatoid arthritis showed many NVA activities, such as recursive steps and lack of crucial information, reducing the time dedicated by the practitioner to each patient. Among the objectives of this second LSS programme were the reduction of non-quality costs and of NVA activities (excessive amount of phone calls, consultations, and requests), besides an increased mutual satisfaction of both patients and health care staff.
In this second case scenario, as well as in the one previously mentioned, the Measure and Analyze steps of process performance singled out that several criticalities were related to a lack of an effective database in which to root the communication exchange among GPs, consultants and nurses on the patients and their medical record. This was one of the main root causes that led to organizational and management issues. Establishing standard internal procedures to exchange information within the health care context under scrutiny and tools like Poka Yoke, the quality of information exchange has increased dramatically.

These examples highlight how a correspondence between optimized health care processes through applied LSS programmes of Continuous Improvement, and the implementation of the same approach to create/improve internal hospital databases – and biomedical databases on an ideal larger scale – would lead to Process Excellence, which would benefit all the participants involved in health care processes. DMAIC guarantees to standardize the research process of the root causes that influence all phenomena. Therefore, selecting and clustering crucial information according to a process that is correct, standardized and quick, is fundamental. Even more for health care, where time and quality assurance are decisive for effective (and timely) patient treatments.

**FINAL REMARKS AND CONCLUSIONS**

The research idea suggested by this article is to foster the application of the LSS methodology as a standard tool for quality assurance of biomedical ontology – specifically, to develop functional health care ontologies meeting the challenge of ensuring reliability of data collection, of creating a common language when speaking with data, and, ultimately, of promoting a shared culture of process optimization and cost effectiveness. Creating LSS-based health care ontologies would implement participative data and information exchange of health care practitioners and LSS consultants, who will be involved in building a corpus of shareable knowledge according to a statistically quantified and optimized procedure that ensures the correct qualitative and quantitative research methods to achieve usable data. Within this scope, integrated applications of LSS, Collaboration Engineering, and Collaboration Ontology will also focus on building ontologies that can be accessible by the different users within health care contexts, who are not necessarily ontology engineers, but could be practitioners or nurses, looking for the fastest and most effective way either to associate symptoms to a specific condition, or to retrieve medical records for a patient and decide treatment. In this article, we presented an attempt to integrate biomedical ontologies and LSS to provide a standardized tool for quality assurance. LSS offers a unique methodology integrating shared language, statistical measuring and effective process management techniques, using DMAIC as versatile tool to provide solutions to interfunctional problems, particularly where people with different competences are involved. The article explored the potential of a standard methodology to process data and information, and for quality assurance of health care ontologies, drawing on Collaboration Engineering and capturing its effective integration with LSS logical model based on the DMAIC cycle. The applications of such integrated approach are likely to be implemented in the most various contexts, which will foster the research debate at both academic and operational levels, thus bringing about new developments in the field.

**REFERENCES**


[31] OWL, https://www.w3.org/OWL/ [last accessed October 20th 2017]


[37] Arcidiacono, G., Giorgetti, A., and Pugliese, M., 2015,
“Axiomatic Design to improve PRM airport assistance,”


