Analysis of Semantically Enriched Process Data for Identifying Process-Biomarkers

Tobias Weller*, Maria Maleshkova*, Martin Wagner†, Lena-Marie Ternes† and Hannes Kenngott†
*Institute of Applied Informatics and Formal Description Methods (KIT), Germany
Email: {tobias.weller,maria.maleshkova}@kit.edu
†Department of General, Visceral and Transplant Surgery, University Hospital Heidelberg, Germany
Email: {Martin.Wagner,Lena-Marie.Ternes,Hannes.Kenngott}@med.uni-heidelberg.de

***Abstract*—Intelligent data analysis is used in multiple domains to find new insights in data that has not been known before. Among others, intelligent data analysis is used in the health-care sector to analyse patient and process data for discovering conclusions and supporting the decison-making process. Besides this, Semantic Web Technologies are currently finding new and broader application areas, including the medical domain. We want to use the advantages of semantic technologies in our data analysis for identifying Process-Biomarkers in medical treatment processes. The semantics allow for a more efficient retrieval of arduous and complex requests, which enables a more intelligent data analysis. This allows for identifying and quantifying effects between different performed tasks and events, in medical treatment processes, in a more detailed way. To address this we 1) extracted information from different data sources; 2) applied Semantic Web Technologies on the extracted information and integrated them into a collaborative platform and enriched them with further semantic background knowledge; 3) performed different statistical methods on the semantically enriched data to identify Process-Biomarkers.**

Keywords–Data Analysis;Healthcare;Semantic Technologies; Semantic MediaWiki

I. INTRODUCTION

Intelligent data analysis uses techniques from Knowledge Discovery and Data Ming to predict the outcome and find patterns in the data. There are approaches available that extend these existing techniques in order to include semantic information about the data [1][2]. This shows among others that Semantic Web Technologies find their way in new application areas, including the medical domain. Hereby, one can use the advantages of Semantic Web Technologies including data integration, data retrieval and intelligent data analysis [3][4]. Semantics enables a machine-readable description of data. Thereby, context information can be included that can be used by machines and in analysis. We want to contribute towards an intelligent data analysis by using Semantic Web Technologies in the medical domain by presenting our approach for identifying Process-Biomarkers.

In the medical domain, Biomarkers [5] are used as indicators to determine a biological condition and to make clinical estimates. Thus Biomarkers are, among others, used to assess the medical condition of an individual. The National Institutes of Health (NIH) defined in 1998 Biological Markers (Biomarkers) as "a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [5].

Although Biomarkers are well-established in the biomedicine, there are no such indicators for medical treatment processes. Therefore, we want to make a first step towards establishing this term as part of a medical treatment process analysis. We introduce in this context "Process-Biomarkers" and use the term as a process related measurement that reflects a coherence between a performed task and event in a medical treatment process.

The considered problem that we want to tackle is to identify and quantify coherences of performed tasks and events in medical treatment processes. We are interested if specific events, like a surgical complication, occurred during a surgery, effects the length of stay of a patient or if the fact that a patient is smoker influences the occurrence of complications during a surgery. Thus, we aim to identify correlations between different parameters in medical treatment processes.

In order to address this problem, we present an approach to identify correlations in medical treatment processes. For this purpose, we exported process and patient data from a Hospital Information System (HIS) and semantified them. We integrated the data with other available data, such as medical treatment flow schema, ontologies to structure the process and other meta-information, into a collaborative platform. All available information in this collaborative platform can be queried and analyzed in order to identify Process-Biomarkers. This semantically enriched data allows not only to analyse the data itself but also include background knowledge in our data analysis.

We demonstrate the applicability of our approach by modeling a concrete medical treatment process and considering information extracted from the Hospital Information System (HIS) for rectal surgeries. Information include habits of the patient, timestamps of different performed tasks, if different complications occurred during the surgery, loss of blood during the surgery and length of stay in the hospital. Overall, we address the following research questions:

1) How can we use different information from multiple sources?
2) How can we implement an infrastructure necessary for storing, accessing, and processing information?
3) Which methods can we apply to identify Process-Biomarkers and enable intelligent data analyis?

The paper is structured as follows. First, we present state of the art in the field of integrating data and applying Semantic Web Technologies on data in section II. In addition, we will show similar approaches to identify correlations in data. Afterwards, we will introduce the goals and the significance of our contributions in Section III. Section IV introduces our proposed methodology to identify correlations in the medical treatment processes and quantify the results. This Section includes the enrichment of the medical data with semantics, handling different scales of measurements in the data and

 ISBN: 978-1-61208-518-0

introducing methods, which we used to identify Process-Biomarkers. We applied multiple methods to identify Process-Biomarkers in order to compare the results and receive multiple views on the process data. After that, we will describe the results of the proposed methodology in Section V. Finally, in Section VI, we will conclude the paper.

II. Related Work

Our approach is addressed by roughly two kinds of work: 1) integrating data from different sources and applying Semantic Web Technologies 2) performing statistical methods to identify correlations in the data.

Previous works covered the aspects of data integration [6][7][8][9]. Hereby, the Biomedical and Live Science domain used Semantic Web Technologies in order to structure and integrate data from different sources [10][11]. Data from different data sources is usually represented in different formats. Therefore, in order to integrate the data from different data sources in one data store, an ETL approach is needed to extract the data and transform them into a common representation. Hereby, approaches that converts the data and structure them in an ontology already exists [12][13]. Usually, the data is only on a local machine available. It is hard for domain experts to contribute to a semantic enrichment. A collaborative approach and providing the possibility to domain experts to enrich the data with their knowledge is preferable. This additional information can be exploited in subsequent data analysis.

For the purpose of capturing and structuring medical treatment processes in BPMN, we had to find suitable ontologies and approaches that allowed to model all considered aspects. There are many ontologies for BPMN 2.0 available [14][15][16] that allows to capture the semantics in processes and include meta-information. However, not all ontologies are online available and follow the latest BPMN 2.0 version. We used a BPMN Ontology from the Data & Knowledge Management (DKM) research unit [17][18] to structure processes. This ontology has a very detailed formalization in OWL 2 DL of the BPMN 2.0 specification. However, we abstract from the ontology and the process modeling, so that the ontology can also be changed in future. This approach enables a flexible representation of knowledge.

Patient related aspects, like, e.g., the type of surgery, were modeled by using OPS Code [19], which provides medical classifications and terminologies for health care systems.

The second aspect that we addressed with our approach is the application of statistical methods on the integrated semantic data in order to identify Process-Biomarkers. In the medical domain are Biomarkers identified by performing experiments, recording the data and applying Student's t-test on the recorded data [20][21][22]. However, we could not perform experiments during a process. Therefore, we used historical data and applied a similar hypothesis test and correlation analysis on the data. Similar approaches were used to identify tumor perfusion related parameters [23], cardiac and metabolic markers in visceral and abdominal subcutaneous [24] and impacts of clinical and genetic features on the clinical progression of Huntington's Disease [25]. In all these approaches could no experiments be performed. They mostly used Spearman's rank correlation to find correlations, however, they did not use other correlation coefficients like, e.g., Kendall in order to compare their results with other statistical methods.

III. Motivation

We want to use Data Science analysis and Semantic Web Technologies and apply them in the medical domain to address our research questions and tackle the raised problems of identifying Process-Biomarkers.

The information, which we will use for Process-Biomarker identification, is distributed across different data sources. In order to use them, we will first integrate the information into one SPARQL store and semantically enrich them with additional information. Afterwards, we can take advantage from the Semantic Web Technologies by efficiently querying the data and easily integrate further information. Our approach to identify Process-Biomarkers is shown in fig. 1. We extracted information such as process data, medical treatment process and patient data. Afterwards, we structured it according to existing ontologies (1). All available information is integrated into a SPARQL store, which was enriched with additional semantic information (2). Afterwards, we queried the data (3) and analyzed it (4) in order to identify Process-Biomarkers.

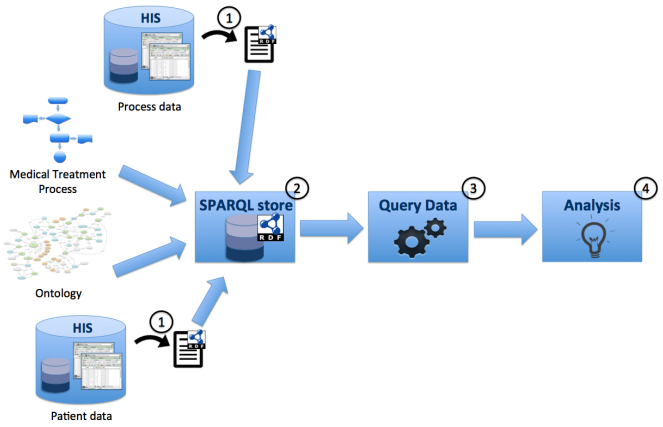

Figure 1. Overview of our proposed approach.

We included semantic information into the SPARQL store, in order to retrieve query results that could not be answered without the new semantic information. Therefore, the knowledge that physicians and nurses are both persons enables to analyse the influence of persons on a process. Retrieving the distinct groups from the raw data sets without semantic is only possible under hard effort. Therefore, the semantic information allow simplified analysis of the data. This kind of knowledge improves the retrieval of data and makes the analysis smarter and therefore more intelligent. The semantic information is machine-readable and can be used by analysis method to comprehend the relations of data for more efficient data analysis.

Another aspect is that the correlations in medical treatment processes are influenced by many different impacts. Thus, the loss of blood during a surgery might depend on the type of surgery, the duration, if complications occurred during the surgery, the age of the patient, and its medical condition. Considering all these impacts in our analysis requires a method to model and query these information.

In addition, the improved retrieval of information allows refined queries. Therefore, we can perform tests for detect-

 ISBN: 978-1-61208-518-0

ing Process-Biomarkers on different abstraction levels on the data. Thus we can use the semantic information and meta-information to refine Process-Biomarkers like e.g the type of surgery does not correlate to the duration of the surgery, but the type of surgery in combination with the surgeon, which performed the surgery, does correlate with the duration of the surgery.

Process-Biomarkers are indicators that help process analysts to comprehend and understand a process and the relations of different process variables and its outcomes in a greater depth. This knowledge is available by the Process-Biomarker, which is a quantified correlation between two performed tasks or events in a medical treatment process. Thus, a process analyst can comprehend the fact that a specific complication during the surgery increases the frequency of blood analysis after the surgery and therefore influences the process. Process-Biomarker also indicates the strength of this coherence. The challenge of finding correlations in processes is the great amount of possible variables might have influences on the process.

In addition, by determining the strength of a coherence, one can state the likelihood and confidence of the Process-Biomarkers. By knowing correlations and confidences of Process-Biomarkers, the prediction of the outcome and the course of a medical treatment process can be improved. Therefore, persons, involved in the medical treatment process, can adapt themselves to possible outcomes and courses. Thus, knowing Process-Biomarkers help process analysts to predict a possible outcome and optimize the likelihood of medical treatment processes to aim a best possible outcome.

Besides predicting future occurrences, Process-Biomarkers can also be used to optimize a process and claim next steps. Therefore, the known correlations and the possible outcome of a medical treatment process can be used to adjust the process for preventing an unwanted outcome. Thus, this knowledge can be used to optimize medical treatment processes according to an established optimization problem.

IV. Semantical Process Correlation Analysis

Our approach is separated into two steps. The first step is the integration of data from multiple sources and the semantic enrichment. This is described in Section IV-A. The second step is the data analysis by exploiting the semantic information, which we integrated into our system. We introduce the methods that we used to identify Process-Biomarkers in Section IV-B.

A. Data Integration and Semantic Enrichment

The first step that we have to tackle is the integration of data from different data sources. We propose a collaborative platform for capturing, annotating and querying process operations, process data, patient data and related process data. Fig. 2 shows a high-level overview of the architecture.

We used a Semantic MediaWiki (SMW) v. 2.3 [26][27] as collaborative platform. This is an extension to MediaWiki [28] that allows for storing information in a structured way and publishes them according to the Linked Data principles [29]. The Semantic MediaWiki performs as the central hub for storing and accessing information. In the backend of the SMW runs an Open Virtuoso Database v. 6.01.3127 that stores the information. Open Virtuoso provides an endpoint that allows

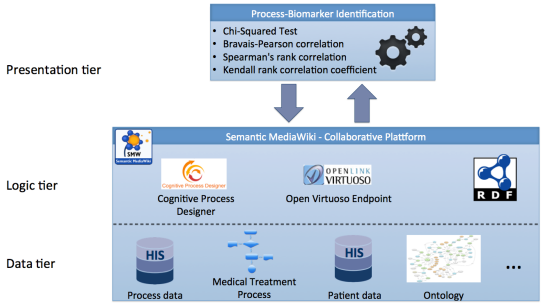

Figure 2. High Level Architecture of the proposed framework.

for querying all available data in the SMW by using SPARQL 1.1 [30]. The advantage of SMW as collaborative platform is the easy access by multiple persons and the easy integration of background knowledge by domain experts through the well-known wiki user interface.

There are two parts of the framework: 1) integrating process data, medical treatment process, patient data and model them with existing ontologies 2) storing and accessing these information semantically in a collaborative platform.

The following data is available for analysis and identifying Process-Biomarkers:

1) Process data from Hospital Information System: The data is stored in a relational database. We exported and structured it according to available ontologies in the medical domain. Available information includes the day of the performed surgery, involved persons, timestamps of each tasks, complications that occurred during the surgery and type of surgery. The data was exported and transformed into RDF.

2) Medical treatment process flows: We modeled the medical treatment process, which shows the sequence of tasks for the considered medical treatment process, by using Business Process Model and Notation (BPMN) as modeling language. The main reason is that BPMN is proposed as a standard by the Object Management Group (OMG) in 2008. The current available version of BPMN is 2.0.2, published in ISO/IEC 19510 [31]. In addition, there are many workflow engines, written in multiple programming languages, available that allows the execution of the processes like, e.g., Oracle Business Process Management Suite [32], Camunda [33] and Sydle [34]. We used an existing ontology from the Data & Knowledge Management (DKM) research unit [17][18]. to structure the information.

3) Patient data from Hospital Information System: This information includes general patient information like gender, weight and age, as well as patient specific information like the habits (e.g. smoker), former diseases (hepatic diseases) and the length of stay in the hospital. It is stored in a relational database. Therefore, we exported and transformed it into a suitable representation according to existing ontologies and standards.

4) Meta-Information: Besides capturing and modeling medical treatment processes and related information, the semantics allow also for including meta-information about the range of values for all parameters in the process. For example, we declared that the parameter *loss of blood* can only take values greater than zero. This knowledge helps to validate data and detect the parameter's scale of measurement.

For the first part of our framework, the process and patient

 ISBN: 978-1-61208-518-0

data was exported from the Hospital Information System in spreadsheet format. Afterwards, the data was uploaded via a programmed bot into the SMW. The medical treatment process was captured with the Cognitive Process Designer [35] v.0.6, which is an extension to SMW that allows for modeling processes in BPMN. The integration of an existing BPMN Ontology [17] occurred manually.

An important aspect for identifying Process-Biomarkers is the comprehensive analysis of the data. Different views on the data allows for different analysis and thus revealing refined statements. Providing semantic information on the available data allows for raising enhanced queries and therefore different views. An example of this circumstance is including the information that wound infection, haemorrhage, abscess and burst abdomen are all complications that might occur after a surgery. By including this information, one can retrieve the information if a complication occurred for a patient after the surgery, without defining each of these complications specifically in a query. The advantage is that we can thus check if a Process-Biomarker exists for wound infection. However, we can also generalize the statement and check if a Process-Biomarker for complications that occur after a surgery exists, like, e.g., if the medical condition of a patient influences the fact of having complications after a surgery. Another advantage is that queries do not have to be adapted if additional information about complications are included in the process data but defining them as complication.

We included process specific information, not given by the ontology, in order to specify the tasks in more detail. These semantic information can later be used, in addition to the available data from the HIS, for analysis and identifying Process-Biomarkers.

We considered different scales of measurements of the data, because the used methods to identify Process-Biomarkers cannot be applied to all scales of measurements. We included the knowledge about the scale of measurement of each parameter by using semantic properties. This knowledge was used later in the analysis step automatically to select the appropriate methods for identifying Process-Biomarkers, as well as validating the data.

B. Data Analysis by Exploiting Semantic Information

We introduced knowledge about the cardinal scape in the semantic properties in the previous section. This knowledge is now used to preprocess the data automatically. For characteristics that follow a cardinal scale, we automatically calculated bins by using Scott's normal reference rule [36]. Assigning the values to bins does, on the one hand, represent a coarsement of the data, however, on the other hand, it simplifies calculations and the accompanying coarsement seems acceptable. Scott's normal reference rule calculates the optimal number of bins. After having the number of bins available, we calculated the width of the bins. Scott's normal reference rule is shown in the following.

$$h = \frac{3.5\sigma}{n^{\frac{1}{3}}}$$

After preprocessing the data, we applied methods for identifying Process-Biomarkers. In statistics, the coherence between two or more variables is calculated by using test statistics and correlation analysis. Different correlation test have different characteristics and advantages. Applying multiple methods on the data allows for comparing and validating the results. Furthermore, not every correlation test can be applied to every scale of measurement. Therefore, we used one test statistic and three correlation coefficients to identify Process-Biomarkers that are presented in the following.

Chi-squared [37] is a statistical hypothesis test that allows for chi square test of independence.

Bravais-Pearson correlation coefficient (classical correlation) [38] computes the correlation between two variables. Its range is between $[-1; 1]$. Whereby, -1 implies a perfectly negative correlation between the variables, zero no correlation and 1 a perfectly positive correlation. A negative correlation leads to an increase of a variable if the respective other decreases and vice versa. A positive correlation of a considered variable leads to an increase of the variable, if the respective other increases and vice versa.

Spearman's rank correlation coefficient [39] is a nonparametric measurement of dependence between two variables. Thus, in contrast to Bravais-Pearson correlation coefficient, it does not assume a linear correlation between the two considered variables and can therefore measure the strength of the correlation between two variables on any monotonic function. The range of the Spearman's rank correlation coefficient is $[-1; 1]$. Whereby, the interpretation of the result is same as for the Bravais-Pearson correlation coefficient.

Kendall rank correlation coefficient [40] is, similar to Spearman's rank correlation coefficient, a nonparametric measurement of dependence between two variables and its range, and interpretation, is the same as for Spearman's rank correlation coefficient.

For identifying Process-Biomarkers, we used Chi-Squared Tests, Bravais-Pearson correlation coefficient, Spearman's rank correlation coefficient and Kendall rank correlation coefficient. We used the Apache Commons Mathematics Library [41] for performing the Chi-Squared Test and calculating the coefficients. By having semantic informations an annotations in the data, we can exploit them in our statistical methods and analysis the data in more detail.

V. EXPERIMENTS

For our experiments, we had 1,690 process instances available that describe an intraoperative medical treatment process of a patient including surgical preparation like, e.g., premedication, type of surgery and related process information like the loss of blood during and complications during the surgery. In average are 72 observations for a patient available and at highest 90 distinct observations.

The data set includes all different scales of measurements. Information like the type of the main surgery, which is described by a nominal scale and information if the patient is smoker (true/false), which is an ordinal scale is available, as well as information about the duration of performed activities and the loss of blood during the surgery, which are both described by a cardinal scale.

We performed on these data chi-squared tests, Bravais-Pearson, Spearman's rank and Kendall rank correlation tests. Due to the semantic, we could query and retrieve semantically related information. For instance if in general there was a

 ISBN: 978-1-61208-518-0

TABLE I. SUBSET OF EXAMPLARY RESULTS FROM THE PROCESS-BIOMARKER IDENTIFICATION.

Variable 1	Variable 2	Chi-Square	Pearson	Spearman	Kendall
Operating Room Block-time	Surgical Complication [y/n]	0.005	0.2061	0.2136	0.1751
Start induction - End removal anesthesia	Grade of anastomotic leak	0.0236	0.1496	0.2435	0.1924
Surgical complication [y/n]	Duration of hospitalisation [d]	0	0.4761	0.5322	0.4453
Incision-Suture	Duration of hospitalisation [d]	0	0.3305	0.3876	0.2764
Operating room-time	Operating Room Block-time	0	0.9708	0.9672	0.8569

complication could be queried because of the information that each specific complication is *typeOf* complication. Likewise, we could also retrieve risk factors by including the information which properties belong to this class. On the one hand, we tested all available data on each other, if the scale of measurements allowed it, and on the other hand, we defined own tests, based on the experience and assumptions of physicians. A script queried the data in the SPARQL store. Afterwards, R [42] was used to perform tests on the retrieved data. This workflow was executed fully automatically.

In total, we performed 2,023 tests on the available data. We performed the tests on the available variables but also on own defined variables like, e.g., the ratio of the surgery by the time of anesthesia and looked if a correlation exist. Due to the amount of correlation test, we show in table I a subset that was evaluated by domain experts.

Physicians evaluated the results according to their knowledge and experience. Some of the results are obviously self-explaining such as that operating room-time and the Operating Room Block time does influence each other. So that the time a patient spends in the operating room block depends on the time he spends in the actual operating room. Other identified Process-Biomarkers are also comprehensible like, e.g., if a surgical complication occurs during the surgery, then the length of stay for a patient in the hospital increases. To comprehend the detected Process-Biomarkers, we visualized the created bins and the number of observations for each bin via a chart. Fig. 3 shows the number of observations in the automatically created classes, separated according to the fact if a patient had a surgical complication or not. It illustrates that the length of stays for a patient indeed is increased, if there was a surgical complication. The query for surgical complication could be performed, because we provided the semantic knowledge of each specific surgical complication into the SMW and therefore could retrieve the corresponding information.

Some other Process-Biomarkers like the Start induction - End removal anesthesia influences the Grade of anastomotic leak is surprisingly, although there is only a small correlation. An anastomotic leak is a lack of tightness of a surgically created hollow organ or vascular anastomosis. The fact that Incision-Suture Time effects the duration of hospitalization is also interesting, but more comprehensible. If the operation takes more time, then it might be a more complicated surgery or complication occurred during the surgery and therefore, the length of stay for a patient increases. Therefore, we suppose that the causality is not effected by the operating time, but by a consequent event.

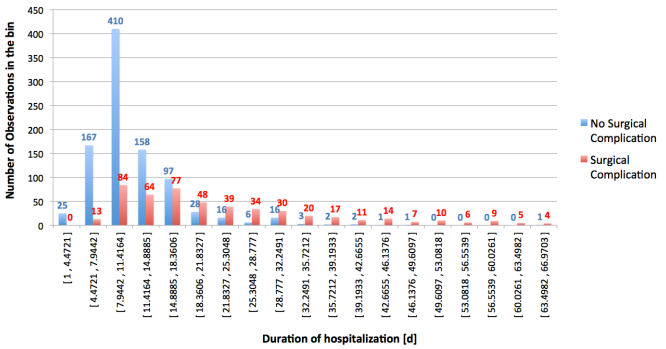

Figure 3. Representation of the bins of the length of stay for a patient, separated if a surgical complication occurred or not.

VI. CONCLUSIONS

We applied semantics in the medical domain in order to contribute towards an intelligent data analysis. We integrated information from different sources into a collaborative platform and used existing concepts and properties from ontologies to describe the medical treatment process and therefore included background knowledge which could be exploited in the data analysis.

The obtained results from the identified Process-Biomarkers help physicians to comprehend effects during the medical treatment process on specific outcome like, e.g., the length of stay for a patient or the loss of blood during a surgery. In addition, the identified Process-Biomarkers help to adjust the medical treatment processes for preventing an unwanted outcome. Besides the improvement of predictions, one can use Process-Biomarkers to optimize processes, in advance, according to specific endpoints such as length of stay for a patient and the mortality. The result of the statistical optimization is an improved process. Still, the Process-Biomarkers can also be used to adapt proceeding processes to prevent adverse results.

Although, much data was available, we want to extend the considered process and the available data with temporal previous and subsequent processes. Thus, we can make statements with identified Process-Biomarkers that go beyond a single process. We focused in this work on the medical domain, however, our approach is generally applicable. Thus, we can apply the used methods in other domains, for instance to identify Process-Biomarkers in business processes.

In conclusion, we have taken a first step towards a statistical analysis of semantically enriched medical treatment data that allows for identifying Process-Biomarkers. The collaborative platform is well-suited for interdisciplinary work and allows for integrating external knowledge. We will show in a future work, that the detected Process-Biomarkers will help us to improve the prediction of values, relevant for the medical treatment process.

 ISBN: 978-1-61208-518-0

REFERENCES

[1] L. Galárraga, C. Teflioudi, K. Hose, and F. M. Suchanek, "Fast rule mining in ontological knowledge bases with amie," The VLDB Journal, vol. 24, no. 6, 2015, pp. 707–730.

[2] C. J. Matheus, G. Piatetsky-shapiro, and D. McNeill, "20 selecting and reporting what is interesting: The kefir application to healthcare data." AAAI Press/MIT Press, 1996.

[3] A. Wiesner, J. Morbach, and W. Marquardt, "Information integration in chemical process engineering based on semantic technologies," Computers & Chemical Engineering, vol. 35, no. 4, 2011, pp. 692–708.

[4] A. Sheth and C. Ramakrishnan, "Semantic (web) technology in action: Ontology driven information systems for search, integration and analysis," IEEE Data Engineering Bulletin, Special issue: Making the Semantic Web Real, vol. 26, 2003, pp. 40–48.

[5] B. D. W. Group, "Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework," Clinical Pharmacology & Therapeutics, vol. 69, no. 3, 2001, pp. 89–95.

[6] S. Jupp and et al., "The ebi rdf platform: Linked open data for the life sciences," Bioinformatics, 2014.

[7] D. Calvanese, P. Liuzzo, A. Mosca, J. Remesal, M. Rezk, and G. Rull, "Ontology-based data integration in epnet: Production and distribution of food during the roman empire," Engineering Applications of Artificial Intelligence, vol. 51, 2016, pp. 212–229, mining the Humanities: Technologies and Applications.

[8] B. Kämpgen, T. Weller, S. O'Riain, C. Weber, and A. Harth, "Accepting the xbrl challenge with linked data for financial data integration," in The Semantic Web: Trends and Challenges: 11th International Conference, ESWC 2014, Anissaras, Greece, May 25-29, 2014. Proceedings. Springer International Publishing, 2014, pp. 595–610.

[9] B. Kämpgen, "Dc proposal: Online analytical processing of statistical linked data," in The Semantic Web – ISWC 2011: 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part II. Springer Berlin, 2011, pp. 301–308.

[10] C. Pasquier, "Biological data integration using semantic web technologies," Biochimie, vol. 90, no. 4, 2008, pp. 584–594, recent advances in complete genome analysis.

[11] T. Katayama and et al., "The 3rd dbcls biohackathon: improving life science data integration with semantic web technologies," Journal of Biomedical Semantics, vol. 4, no. 1, 2013, pp. 1–17.

[12] M. Niinimäki and T. Niemi, "An etl process for olap using rdf/owl ontologies," in Journal on Data Semantics XIII. Springer Berlin, 2009, pp. 97–119.

[13] V. Nebot, R. Berlanga, J. M. Pérez, M. J. Aramburu, and T. B. Pedersen, "Multidimensional integrated ontologies: A framework for designing semantic data warehouses," in Journal on Data Semantics XIII. Springer Berlin, 2009, pp. 1–36.

[14] C. Natschläger, "Towards a bpmn 2.0 ontology," in Business Process Model and Notation: Third International Workshop, BPMN 2011, Lucerne, Switzerland, November 21-22, 2011. Proceedings. Springer Berlin, 2011, pp. 1–15.

[15] J. vom Brocke and M. Rosemann, Eds., BPMN 2.0 for Modeling Business Processes, Handbook on Business Process Management 1: Introduction, Methods, and Information Systems. Springer, 2015, ISBN: 978-3-642-45099-0.

[16] W. Yao and A. Kumar, "Conflexflow: Integrating flexible clinical pathways into clinical decision support systems using context and rules," Decision Support Systems, vol. 55, no. 2, 2013, pp. 499–515, 1. Analytics and Modeling for Better HealthCare 2. Decision Making in Healthcare.

[17] M. Rospocher, C. Ghidini, and L. Serafini, "An ontology for the business process modelling notation formal ontology," in Information Systems – Proceedings of the Eighth International Conference. IOS PRess BV, Sep. 2014, pp. 133–146.

[18] "Data & knowledge management," URL: https://dkm.fbk.eu [accessed: 2016-10-04].

[19] "German institute of medical documentation and information," URL: https://www.dimdi.de/static/en [accessed: 2016-10-04].

[20] M.-C. van de Beek and et al., "C26:0-carnitine is a new biomarker for x-linked adrenoleukodystrophy in mice and man," PLoS ONE, vol. 11, no. 4, April 2016, pp. 1–19.

[21] S. Huang and et al., "Attenuation of microrna-16 derepresses the cyclins d1, d2 and e1 to provoke cardiomyocyte hypertrophy," Journal of Cellular and Molecular Medicine, vol. 19, no. 3, 2015, pp. 608–619.

[22] J. L. et al., "Pkc interacts with {STAT3} and promotes its activation in cardiomyocyte hypertrophy," Journal of Pharmacological Sciences, vol. 132, no. 1, 2016, pp. 15–23.

[23] H.-J. Lee and et al., "Tumor perfusion-related parameter of diffusion-weighted magnetic resonance imaging: Correlation with histological microvessel density," Magnetic Resonance in Medicine, vol. 71, no. 4, 2014, pp. 1554–1558.

[24] I. J. Neeland and et al., "Associations of visceral and abdominal subcutaneous adipose tissue with markers of cardiac and metabolic risk in obese adults," Obesity, vol. 21, no. 9, 2013, pp. E439–E447.

[25] I. Dogan and et al., "Consistent neurodegeneration and its association with clinical progression in huntington's disease: A coordinate-based meta-analysis," Neurodegenerative Diseases, vol. 12, no. 1, 2013, pp. 23–35.

[26] M. Krötzsch, D. Vrandečić, and M. Völkel, "Semantic mediawiki," in The Semantic Web - ISWC 2006: 5th International Semantic Web Conference, ISWC 2006, Athens, USA, November 5-9, 2006. Proceedings. Springer Berlin, 2006, pp. 935–942.

[27] "Semantic mediawiki project," URL: https://www.semantic-mediawiki.org [accessed: 2016-10-04].

[28] "Mediawiki," URL: https://www.mediawiki.org [accessed: 2016-10-04].

[29] "Linked data," URL: https://www.w3.org/DesignIssues/LinkedData.html [accessed: 2016-10-04].

[30] "Sparql," URL: https://www.w3.org/TR/sparql11-overview/ [accessed: 2016-10-04].

[31] "Information technology – object management group business process model and notation," URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=62652 [accessed: 2016-10-04].

[32] "Oracle," URL: http://www.oracle.com/us/technologies/bpm/suite/overview/index.html [accessed: 2016-10-04].

[33] "Camunda Services GmbH," URL: https://camunda.org [accessed: 2016-10-04].

[34] "Sydle," URL: http://www.sydle.com/en/bpms/ [accessed: 2016-10-04].

[35] T. Weller and M. Maleshkova, "Capturing and annotating processes using a collaborative platform," in Proceedings of the 25th International Conference Companion on World Wide Web, ser. WWW '16 Companion. International World Wide Web Conferences Steering Committee, 2016, pp. 283–284.

[36] D. W. Scott, "On optimal and data-based histograms," Biometrika, vol. 66, no. 3, 1979, pp. 605–610.

[37] K. Pearson, "On the criterion that a given system of derivations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," Philosophical Magazine, vol. 50, no. 302, 1900, pp. 157–175.

[38] K. Spearman, "Note on regression and inheritance in the case of two parents," Proceedings of the Royal Society of London, vol. 58, no. 347-352, 1895, pp. 240–242.

[39] C. Spearman, "The proof and measurement of association between two things," The American Journal of Psychology, vol. 15, no. 1, 1904, pp. 72–101.

[40] M. G. Kendall, "A new measure of rank correlation," Biometrika, vol. 30, no. 1/2, 1938, pp. 81–93.

[41] "Commons math: The apache commons mathematics library," URL: https://commons.apache.org/proper/commons-math/ [accessed: 2016-10-04].

[42] "The r project," URL: https://www.r-project.org/ [accessed: 2016-10-04].

 ISBN: 978-1-61208-518-0