

Arden Syntax MLM Building Blocks for Microbiological Concepts and Their Application in Infection Surveillance

Jeroen S. DE BRUIN^{a,b,1}, Walter KOLLER^c, Julia ZECKL^a, Alexander BLACKY^d, Andrea RAPPELSBERGER^b, and Klaus-Peter ADLASSNIG^{a,b}

^a Medexter Healthcare GmbH, Vienna, Austria

^b Section for Artificial Intelligence and Decision Support, CeMSIIS, Medical University of Vienna, Austria

^c Department of Hospital Epidemiology and Infection Control, Medical University of Vienna and Vienna General Hospital, Austria

^d VAMED-KMB Hospital Management and Operation GmbH, Vienna, Austria

Abstract. *Background:* The diagnosis – and hence definitions – of healthcare-associated infections (HAIs) rely on microbiological laboratory test results in specific constellations. *Objectives:* To construct a library that provides interoperable building blocks for the analysis of microbiological laboratory test results. *Methods:* We used Java for preprocessing raw microbiological laboratory test results and Arden Syntax for knowledge-based querying of data based on microbiology information elements used in European surveillance criteria for HAIs. To test the library and quantify how often these information elements occur in the data, we performed a retrospective cohort study on adult patients admitted for at least 24 hours to an intensive care unit at the Vienna General Hospital in 2013. *Results:* We identified eleven information elements for which information was electronically available. These elements were identified positively 1,239 times in 1,184 positive microbiology tests from 563 patients. *Discussion:* The availability of a library for the analysis of microbiology laboratory test results in HAI terms facilitates electronic HAI surveillance.

Keywords. Clinical decision support, medical knowledge bases, Arden Syntax, MLM building blocks, infection control and surveillance.

1. Introduction

More than a half of all healthcare-associated infections (HAIs) – i.e., infections occurring in a patient during treatment in a hospital or other healthcare facility [1] – are of bacterial or fungal origin [2]. Based on pathogen identification and antibiotic susceptibility testing, the microbiological laboratory helps to verify a suspected infection and provides crucial information for selecting the appropriate infection treatment. This is especially important in intensive care units (ICUs), where the patient's health is greatly compromised and the prevalence of multi-resistant organisms resistive to standard antibiotic therapies is higher [3, 4].

International infection surveillance programs, which include European surveillance definitions of HAIs in ICUs by the European Centre for Disease Prevention and Control

¹ Corresponding Author: Jeroen S. de Bruin, Medexter Healthcare GmbH, Borschkegasse 7/5, 1090 Vienna, Austria, E-Mail: jb@medexter.com

(ECDC) in Stockholm [5] require microbiology results to verify most types of HAIs. This information includes the type of sample material taken from the patient; the type of tests performed on the sample; the type of microorganism detected, the pathogen's abundance in the tested sample, and the time of detection.

The digitization of microbiological laboratory test results facilitates their rapid communication and permits various modes of automatic processing, such as the computerized detection of a bacterial outbreak or infection [6–9]. An essential prerequisite for computerized use of microbiological laboratory test results is their storage and communication in a structured and standardized format.

Our goal was to create a standardized library for the analysis of microbiological laboratory test results. The library is intended to provide interoperable building blocks for the analysis of microbiology results in knowledge-based infection detection and surveillance systems. Building blocks correspond with microbiological information elements that are frequently required by HAI surveillance programs, such as the aforementioned ECDC infection surveillance criteria. Preprocessing and encoding of raw clinical data were implemented in Java. Subsequent data querying for the aforementioned information elements was implemented in Arden Syntax, an international standard for computerized knowledge representation and processing that supports the collection, description, and processing of medical knowledge in a machine-executable format [10]. To test the library and assess how frequent implemented queries were performed, we focused on microbiology data from ICUs of the Vienna General Hospital (VGH).

2. Methods

2.1. Study Design, Setting, and Participants

We conducted a retrospective cohort study on prospectively collected and validated data. The study was performed at VGH, a 1,922-bed tertiary-care and teaching hospital. All adult patients (age ≥ 18 years) who were admitted to one of the VGH's ICUs for at least 24 hours between 1 January and 31 December 2013, and for whom positive microbiological test results were available, were eligible for the study.

2.2. Data Acquisition and Management

Microbiological laboratory test results were acquired from the MOLIS laboratory information system at the Department of Clinical Microbiology at VGH. Data are structured in XML format and include unique patient and sample identifiers, sample sender (department), demographic patient data (age, sex), the number of samples tested, sample source, performed tests, and the pathogens that tested positive.

Using unique patient identifiers, the MOLIS database was cross-referenced with the KisDB hospital information system database of VGH. SQL queries were used to obtain information on a patient's hospital stay, including the unit(s) the patient stayed in during the present hospital stay, when the patient was admitted to the unit, and for how long.

Data acquired from both sources were combined and stored in the MiBiDB project database, which is an Oracle relational database.

2.3. Knowledge Base Implementation and Processing

The library was implemented using Java and Arden Syntax. Data preprocessing and preparation was implemented in Java.

Data preprocessing and preparation were done using a custom-made thesaurus of bacteria, with which microbiology results could be classified according to bacteria type (such as a common skin contaminant, a uropathogen, etc.), and according to sample site (blood culture, catheter culture). The structure of the thesaurus was based on the NCBI taxonomy browser [11], and was initially filled with and coded according to the microorganisms in the code list provided in [5]. The thesaurus was then extended to include the analysis of about 450.000 MOLIS XML files, from which we extracted all pathogens found. The extracted data included at least pathogen family, gender, species, and name as stated in the XML file. The pathogens were then classified manually by microbiology experts.

When processing new microbiology results, matching algorithms analyze both the coded input in the XML files and the pathogen name – which is provided by the microbiology laboratory technician as a free text input. Through this process, reported pathogens are matched to any of the bacteria present in the thesaurus and subsequently assigned the class(es) of the corresponding set(s) of the respective bacterial strain.

The building blocks containing rules for information elements were encoded in Arden Syntax. With Arden Syntax, medical rules are coded in a syntax that resembles natural language, which makes the code more easily comprehensible and verifiable by healthcare professionals [12]. The medical rule sets are known as medical logic modules (MLMs), and usually contain sufficient logic to make at least a single medical decision [13].

For coding and processing Arden Syntax MLMs, we used the ArdenSuite framework [14], which includes an integrated development environment (IDE) as well as an ArdenSuite server. The IDE was employed for coding, compiling, and testing MLMs before transferring them to the ArdenSuite server, where compiled MLMs are stored and executed. For server access, the ArdenSuite server uses web-service protocols: either the Simple Object Access Protocol (SOAP) or Representational State Transfer (REST). The server/database connector employs Java Database Connectivity (JDBC); this additional module may be used to connect the ArdenSuite server with any SQL-based external database sources. For the present project we used the server/database connector to connect with the MiBiDB project database. Figure 1 provides a graphic diagram of the project architecture.

2.4. Presentation of Results

We used patient demographics (age, sex, length of stay) to describe the patient population. For the identification of microbiological information needed for infection allocation, we analyzed ECDC infection surveillance criteria for ICUs, and presented those information elements (see Table 1). Elements are numbered and grouped according to their related HAI syndrome(s). Based on this analysis and the analysis of available microbiology data, information elements were implemented in the library. They are referenced by their related HAI syndrome and – if more than one description exists – by their respective

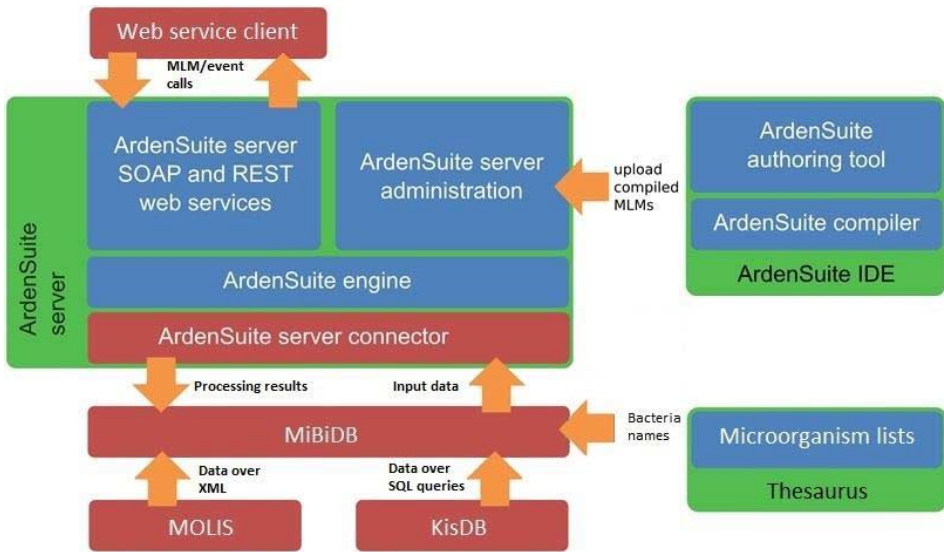


Figure 1. The ArdenSuite framework for medical knowledge representation and rule-based inference. Image adapted from [14]. Note: MLM, medical logic module.

sub-category number (such as PN3.3). These are presented in Table 2, along with the frequency (both absolute and relative to the total number of positive results found) of these library elements in the study data.

3. Results

3.1. Identification of Microbiological Information Elements

Table 1 shows the results of the analysis for ECDC infection surveillance criteria in ICUs. In all, 28 information elements related to microbiology results were found.

3.2. Library Analysis

For each of the 28 identified microbiological information elements, an Arden Syntax MLM was written by VGH infection control and microbiology experts together with medical knowledge engineers. After applying these MLMs to the study data in the MiBiDB database, eleven of these MLMs yielded one or more positive results.

In total, 1,239 positively identified information elements in 1,184 positive microbiology test samples taken from 563 patients. Thirty-nine percent of the patients were female. The patients' mean age was 59 years (minimum 18; maximum 92; interquartile range 22), and the average length of their hospital stay 20 days (minimum 1; maximum 136; interquartile range 18). Table 2 shows the eleven identified information elements and their absolute and relative frequencies in the study data.

To clarify the workings of the library, consider Figure 2. We outlined how the library handles information elements that involve common skin contaminants or common uropathogens. In the first step, data are transferred from KisDB and MOLIS. In the

Table 1. 28 microbiological information elements for the ECDC ICU infection surveillance criteria.

Information element (related HAI syndrome ²)	Description ¹
<i>UO</i> : Blood only positive for a pathogen in at least one blood culture	Evidence of pathogens in the bloodstream but not present in specimens taken within a reasonable period of time from other potential infection sites. Pathogens include microorganisms not considered to be common skin contaminants.
<i>Sec BSI</i> : Blood and another specimen positive for a pathogen	Evidence of the same pathogen in the bloodstream as well as in another specimen taken promptly. The location of the other specimen determines which HAI syndrome applies. Pathogens include microorganisms not considered to be common skin contaminants.
<i>former BSI-A</i> : Blood positive for the same common skin contaminant in two separate blood cultures	The same common skin contaminant is present in (at least) two blood cultures taken within 48 hours.
<i>PN1</i> : Positive quantitative culture from a minimally contaminated LRT specimen	1. BAL with a threshold of $\geq 10^4$ CFU/ml or $\geq 5\%$ of BAL-obtained cells contain intracellular bacteria on direct microscopic exam (classified on the diagnostic category BAL). 2. Protected brush with a threshold of $\geq 10^3$ CFU/ml 3. Distal protected aspirate with a threshold of $\geq 10^3$ CFU/ml
<i>PN2</i> : Positive quantitative culture from a possibly contaminated LRT specimen	Quantitative culture of LRT specimen (such as an endotracheal aspirate) $\geq 10^6$ CFU/ml
<i>PN3</i> : Alternative microbiology methods indicative of respiratory tract infection	1. Positive blood culture not related to another source of infection 2. Positive growth in culture of pleural fluid 3. Pleural or pulmonary abscess with positive (culture, antigen detection) needle aspiration 4. Histologic pulmonary exam shows evidence of pneumonia. 5. Positive exams for pneumonia with virus or particular germs (<i>Legionella</i> , <i>Aspergillus</i> , mycobacteria, mycoplasma, <i>Pneumocystis carinii</i>) 6. Positive detection of viral antigen or antibody from respiratory secretions 7. Positive direct exam or positive culture from tissue 8. Seroconversion (example: influenza viruses, <i>Legionella</i> , <i>Chlamydia</i>) 9. Detection of antigens in urine (<i>Legionella</i>)
<i>PN4</i> : Positive sputum	1. Positive sputum culture 2. Positive non-quantitative LRT specimen culture
<i>UTI-A</i> : Positive urine culture	Urine culture with $\geq 10^5$ CFU/ml of at most two different microorganisms
<i>UTI-B</i> : Conditional positive urine culture	1. At least two urine cultures with repeated isolation of the same uropathogen (GNB or <i>S. saprophyticus</i>) with $\geq 10^2$ colonies/ml urine in non-voided specimens 2. $\leq 10^5$ colonies/ml of a single uropathogen (GNB or <i>S. saprophyticus</i>) in a patient being treated with an effective antimicrobial agent for a urinary infection
<i>CRI3</i> : Positive catheter culture	1. Quantitative culture of a catheter tip ($\geq 10^3$ CFU / ml) 2. Semiquantitative culture of a catheter tip (> 15 CFU / ml) 3. Quantitative positive culture of blood drawn from a central venous catheter; with CFU/ml reported ³) 4. Quantitative positive culture of blood drawn from a peripheral vein; with CFU/ml reported ³) 5. Qualitative positive culture of blood drawn from a central venous catheter ⁴) 6. Qualitative positive culture of blood drawn from a peripheral vein ⁴) 7. Positive culture of pus from the insertion site

Note: HAI, healthcare-associated infection; BAL, bronchoalveolar lavage; CFU, colony forming unit; GNB, Gram-negative bacteria; LRT, lower respiratory tract.

¹) Refer to [5]

²) HAI syndrome explanations: UO, bloodstream infection of unknown origin; BSI, bloodstream infection; sec BSI, BSI secondary to another infection site; PN, pneumonia; UTI, urinary tract infection; CRI, catheter-related infection.

³) Specimens taken and tested in parallel; CFU ratio between catheter and peripheral sample must be > 5

⁴) Specimens taken and tested in parallel; catheter sample must test positive at least 2 hours before withdrawal of the peripheral blood sample.

Table 2. Results for the Arden Syntax library for the eleven identified information elements.

Bloodstream infections (BSI)			Pneumoniae (PN)		
Element	Freq. (abs)	Freq. (rel)	Element	Freq. (abs)	Freq. (rel)
UO	82	6.62%	PN1.1	432	34.87%
Sec BSI	46	3.71%	PN3.2	23	1.86%
BSI-A	5	0.40%	PN4.1	18	1.45%
			PN4.2	1	0.08%

Urinary tract infections (UTI)			Catheter-related infections (CRI)		
Element	Freq. (abs)	Freq. (rel)	Element	Freq. (abs)	Freq. (rel)
UTI-A	359	28.98%	CRI3.1	1	0.08%
UTI-B.2	29	2.34%	CRI3.2	243	19.61%

Note: UO, bloodstream infection of unknown origin; Sec BSI, bloodstream infection secondary to another infection site.

second step, preprocessing algorithms identify and classify bacteria according to the thesaurus. Finally, based on these classifications and additional information such as the pathogen’s abundance in the tested sample and the time of detection, the MLMs test for occurrence of their respective information element.

4. Discussion

In the present report, we discuss the results of a knowledge-based program library for the analysis of microbiology test results for infection control. Using Java for preprocessing raw microbiology test results and Arden Syntax for knowledge-based querying of preprocessed data, we constructed a collection of 28 building blocks that process microbiological information required in the ECDC infection surveillance criteria for ICUs. We then used data from ICUs of the VGH to test the library and quantify how often these information elements appear in the data.

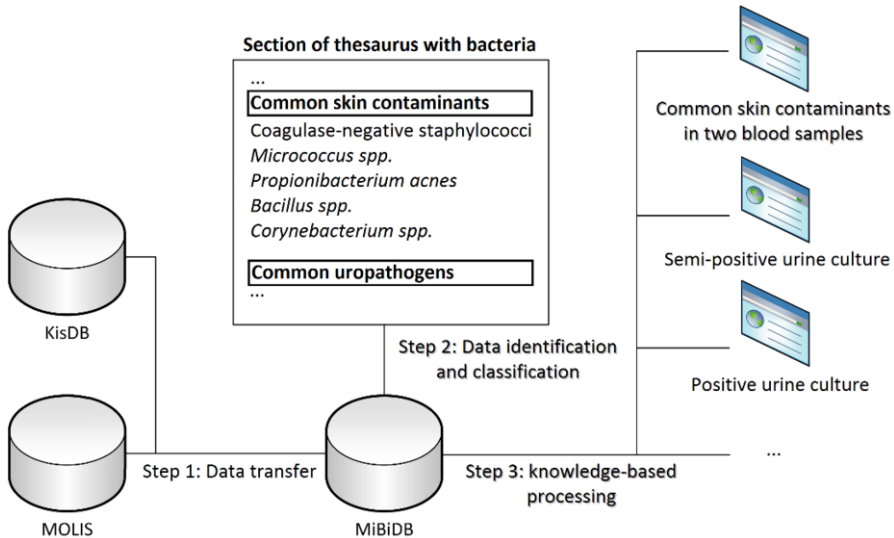


Figure 2. Graphic representation of the identification of microbiological information elements in ECDC infection surveillance criteria for ICUs. Information elements tested for are *common skin contaminants in two blood cultures* (BSI-A), *semi-positive urine culture* (UTI-B.2) and *positive urine culture* (UTI-A).

Infection control remains an important issue in healthcare institutions. Early studies have shown that infection surveillance and control programs in hospitals may reduce the number of HAIs by as much as 32% [15, 16]. However, these programs impose a considerable burden on hospital resources as long as the identification of HAIs depends entirely on the footwork and brainwork of human experts [17]. Electronic infection detection and surveillance could alleviate those burdens. The availability of a standardized knowledge-based library for the analysis of microbiology test results in HAI terms facilitates electronic HAI surveillance, which in turn releases human resources for infection control measures at the bedside.

A central aspect of our method is matching pathogens with a thesaurus. Although the thesaurus is based on the list of bacteria provided in [5], the classification of those microorganisms into sublists was based on the knowledge and experience of infection control experts. Furthermore, although matching algorithms are available to compare microbiology results with thesaurus entries, the thesaurus is manually extended and maintained to capture and compensate for spelling variations of pathogens and record new ones.

Analysis of the study data revealed that the information elements appear in various frequencies. Straightforward and less comprehensive tests, such as positive urine cultures or positive bronchial lavages yielded positive results more often, while restrictive tests such as quantitative catheter cultures or common skin contaminants in two blood cultures rarely appeared in the data.

The limitations of the study are worthy of note. We performed the investigation at a single center. As such, we cannot make claims about the applicability or consistency of the results with respect to healthcare institutions other than VGH. Furthermore, the library is currently based only on surveillance criteria by the ECDC. Other widely used surveillance criteria, such as those of the Centers for Disease Control and Prevention in Atlanta or the Robert Koch Institute in Berlin were not included in the study.

Our current goals are to validate the library and optimize it for the ECDC infection surveillance criteria before we include other surveillance programs. Simultaneously, we are working on a more standardized way of communicating microbiology results. Various communication protocols are available, such as IHE cross-document sharing (XSD), HL7 Version 3 normative edition – Laboratory Domain, HL7 Observation Result messages, or Fast Healthcare Interoperability Resources (FHIR) Diagnosis Report resources. Standardized communication is also related to the standardization of report contents using ontologies such as SNOMED CT and its incorporated (laboratory-specific) Logical Observation Identifiers Names and Codes (LOINC) Ontology. Currently we are also assessing the adoption of ontologies using the ELGA terminology server [18].

References

- [1] World Health Organization, The Burden of Health Care-Associated Infection Worldwide, 2017. Available at: http://www.who.int/gpsc/country_work/burden_hcai/en/, last access: 27.1.2017.
- [2] European Centre for Disease Prevention and Control (ECDC), Point Prevalence Survey of Healthcare-Associated Infections and Antimicrobial Use in European Acute Care Hospitals 2011–2012, 2013. Available at: <http://ecdc.europa.eu/en/publications/Publications/healthcare-associated-infections-antimicrobial-use-PPS.pdf>, last access: 16.11.2016.
- [3] N. Brusselsaers, D. Vogelaers, S. Blot, The rising problem of antimicrobial resistance in the intensive care unit, *Annals of Intensive Care* 1 (2011), 47; doi:10.1186/2110-5820-1-47.

- [4] G. Zilahi, A. Artigas, I. Martin-Loeches, What's new in multidrug-resistant pathogens in the ICU?, *Annals of Intensive Care* 6 (2016), 96; doi: 10.1186/s13613-016-0199-4.
- [5] European Centre for Disease Prevention and Control (ECDC), European Surveillance of Healthcare-Associated Infections in Intensive Care Units, HAI-Net ICU Protocol, Protocol Version v1.02, 2015. Available at: <http://ecdc.europa.eu/en/publications/Publications/healthcare-associated-infections-HAI-ICU-protocol.pdf>, last access: 16.1.2017.
- [6] R.B. Dessau, P. Steenberg, Computerized surveillance in clinical microbiology with time series analysis, *Journal of Clinical Microbiology* 31(4) (1993), 857–860.
- [7] D.M. Hacek, R.L. Cordell, G.A. Noskin, L.R. Peterson, Computer-assisted surveillance for detecting clonal outbreaks of nosocomial infection, *Journal of Clinical Microbiology* 42(3) (2004), 1170–1175.
- [8] J.S. de Bruin, W. Seeling, C. Schuh, Data use and effectiveness in electronic surveillance of healthcare associated infections in the 21st century: a systematic review, *Journal of the American Medical Informatics Association* 21(5) (2014), 942–951.
- [9] R. Freeman, L.S.P. Moore, L. García Álvarez, A. Charlett, A. Holmes, Advances in electronic surveillance for healthcare-associated infections in the 21st Century: a systematic review, *The Journal of Hospital Infection* 84(2) (2013), 106–119.
- [10] Health Level Seven, Arden Syntax v2.10 (Health Level Seven Arden Syntax for Medical Logic Systems, Version 2.10), 2015. Available at: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=372, last access: 30.1.2017.
- [11] E.W. Sayers, T. Barrett, D.A. Benson, S.H. Bryant, K. Canese, V. Chetvermin, D.M. Church, M. DiCuccio, R. Edgar, S. Federhen, M. Feolo, L.Y. Geer, W. Helmberg, Y. Kapustin, D. Landsman, D.J. Lipman, T.L. Madden, D.R. Maglott, V. Miller, I. Mizrahi, J. Ostell, K.D. Pruitt, G.D. Schuler, E. Sequeira, S.T. Sherry, M. Shumway, K. Sirotkin, A. Souvorov, G. Starchenko, T.A. Tatusova, L. Wagner, E. Yaschenko, J. Ye, Database resources of the National Center for Biotechnology Information, *Nucleic Acids Research* 37 (2009), D5–D15.
- [12] M. Samwald, K. Fehre, J. de Bruin, K.-P. Adlassnig, The Arden Syntax standard for clinical decision support: Experiences and directions, *Journal of Biomedical Informatics* 45(4) (2012), 711–718.
- [13] G. Hripsak, Writing Arden Syntax Medical Logic Modules, *Computers in Biology and Medicine* 24(5) (1994), 331–363.
- [14] K.-P. Adlassnig, K. Fehre, Service-oriented Fuzzy-Arden-Syntax-based clinical decision support, *Indian Journal of Medical Informatics* 8(2) (2014), 75–79.
- [15] D.H. Culver, T.C. Horan, R.P. Gaynes, W.J. Martone, W.R. Jarvis, T.G. Emori, S.N. Banerjee, J.R. Edwards, J.S. Tolson, T.S. Henderson, J.M. Hughes, and the National Nosocomial Infections Surveillance System, Atlanta, Georgia, Surgical wound infection rates by wound class, operative procedure, and patient risk index, *The American Journal of Medicine* 91(3B) (1991), 152S–157S.
- [16] R.W. Haley, D.H. Culver, J.W. White, W.M. Morgan, T.G. Emori, V.P. Munn, T.M. Hooton, The efficacy of infection surveillance and control programs in preventing nosocomial infections in US hospitals, *American Journal of Epidemiology* 121(2) (1985), 182–205.
- [17] M. Klompas, D.S. Yokoe, Automated surveillance of health care-associated infections, *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 48(9) (2009), 1268–1275.
- [18] BMG Österreich, ELGA GmbH, FH Technikum Wien, BRZ Österreich, Fachhochschule Dortmund, Terminologie-Browser. Available at: <https://termpub.gesundheit.gv.at/TermBrowser/gui/main/main.zul>, last access: 5.3.2017.