MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190425

Augmenting Analytics Software for Clinical Microbiology by Man-Machine Interaction

Walter Koller[a], Gabriel Kleinoscheg[b], Birgit Willinger[c], Andrea Rappelsberger[d], Klaus-Peter Adlassnig[b,d]

[a] *Department for Hospital Epidemiology and Infection Control, Vienna General Hospital and Medical University of Vienna, Vienna, Austria*
[b] *Medexter Healthcare GmbH, Vienna, Austria*
[c] *Division of Clinical Microbiology, Vienna General Hospital and Medical University of Vienna, Vienna, Austria*
[d] *Section for Artificial Intelligence and Decision Support, Medical University of Vienna, Vienna, Austria*

Abstract

In the present study, we intended to solve identification problems in analyzing the results of microbiology by proactive man-machine interaction. We modified the analytics software MOMO so that it flags laboratory results containing textual elements unknown to the thesaurus, and a human expert assigns the elements to the respective existing thesaurus elements or creates new ones. In 773,309 laboratory results, roughly 2.6% contained unassigned elements and would have been ignored in thesaurus-based analyses for purposes other than simply reporting microbiological findings to physicians. In current use, the thesaurus is kept up to date with synonyms, syntactic deviations, misspellings, and entries not contained earlier, with man-machine interaction of 2–3 hours per week. This approach helps to accommodate both up-to-date clinical reporting for immediate patient care as well as up-to-date queries for infection surveillance and epidemiology, outbreak management, quality control and benchmarking, and antimicrobial stewardship.

Keywords:
Data Analytics; Software; Microbiology

Introduction

Why are clinical information technology (IT) solutions – despite high sophistication and the latest IT standards [1] – sometimes not well accepted by users? This question has been addressed in extensive clinical informatics studies [e.g., 2-4] as well as studies focused on our topic [5, 6], and is still impeding the optimal use of IT tools in clinical and laboratory medicine. A common feature of all studies cited above is the attempt to overcome shortcomings in daily routine outputs which could be allocated to insufficient involvement of human experts during the execution of IT tasks.

In the course of our long-standing involvement in fully automated IT-assisted analysis of microbiological and clinical data, we learned that strict adherence to coded data alone is not enough to avoid deficits in IT analysis [7]. We looked into the discrepancies that emerged when comparing the results of automated IT analyses with the respective gold standards. The search for the root cause led us to elements and terms in our clinical reporting scheme which were not allocated to the thesaurus (terminology coding of the laboratory information system (LIS)), and therefore could not be recognized by our software. In many instances, simple misspellings or orthographic variants were the cause. In other cases, a number, or even whole arrays, of different entities were allocated to a single code, which is why a distinction by code was no longer feasible. So far, such deficits – provided they are detected at all – could explain mistrust in automated IT tools and call for a scrupulous check of each data entry and manual thesaurus allocation of missing terms by human experts. This hardship may be accepted for a research study but is not a realistic approach for a reliable routine IT clinical reporting tool.

Two aspects should be mentioned here: First, free-text entries have been introduced by the users of the microbiology LIS (with arguments discussed later in this paper), thus "invading" the LIS which originally had been focused on merely coded entries and results. Second, microbiology findings play an important role not only in immediate patient treatment but also in contributory disciplines, such as infection surveillance, outbreak management, and antimicrobial stewardship. Hence, microbiology reports and the respective meta-analyses/queries are significant and must therefore be concise, correct, and – last but not least – timely.

The aim of this study was to enhance the precision of our automated analytics and clinical software by adding and augmenting man-machine interaction. In detail, we aimed to solve ambiguity and identification problems in digital reports of clinical microbiology by proactive man-machine intervention.

Methods

General

We modified our automated analytics software so that it flagged laboratory reports which contained elements unknown to the thesaurus and forwarded these to a human expert on a regular basis. The expert then assigned the textual elements to the respective existing entities or created new ones. Thus, we "trained" the thesaurus on a regular basis to recognize possible synonyms, syntactic deviations and misspellings, and thus be extended by new entries.

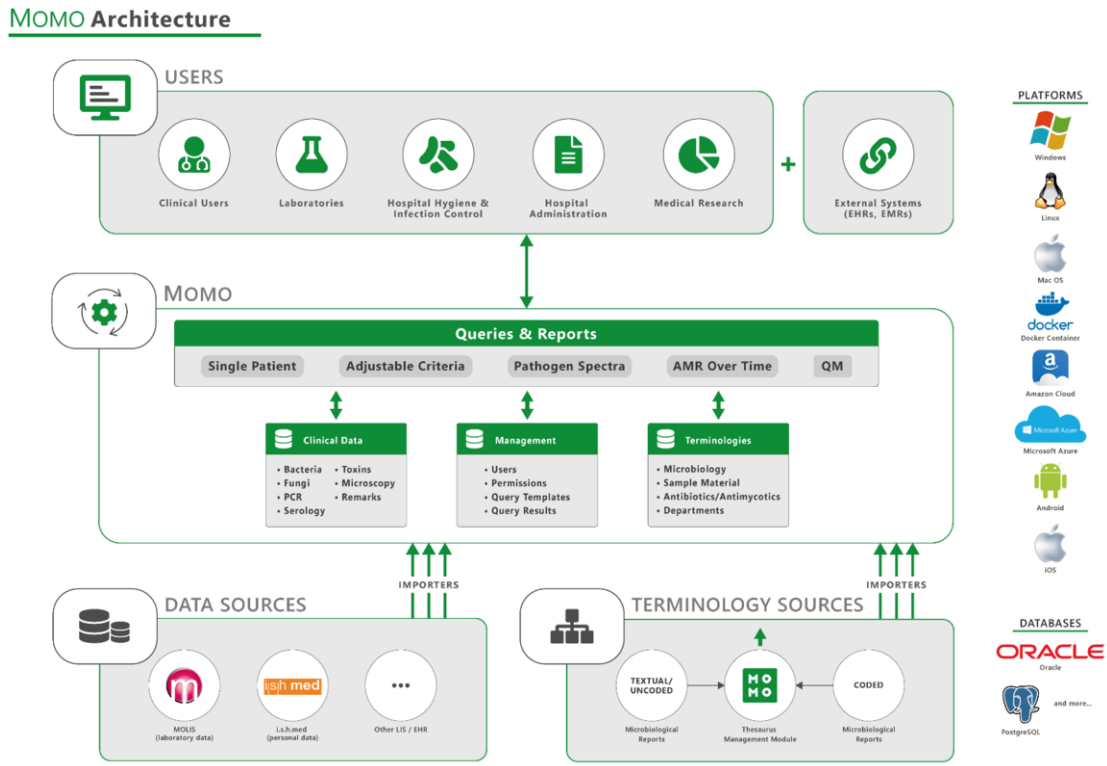

Figure 1 – MOMO is a multifunctional tool for analyzing, monitoring, and reporting pathogens and antimicrobial resistance. It receives 58 different parameters from the microbiology laboratory; four of them are based on terminologies. Most data are structured or coded, some are textual. The latter include not only several comments or report additions, but also microbiological terminology such as bacteria, fungi, PCR, serology, toxins, and microscopy.

Study Setting and Design

We performed a retrospective single-center analysis of validated clinical microbiology results from Vienna General Hospital (VGH), Austria – a 1,900-bed tertiary-care and teaching hospital. Laboratory data were obtained through systematic interrogation of MOLIS (Modular Open Laboratory Information System, Compu Group Medical (CGM) LAB Belgium S.A., Barchon, Belgium) [8] from patients of all VGH clinics from July 4, 2013 to February 16, 2018. MOLIS is designed for IT support of laboratory processes and issues laboratory findings in digital reports, which are delivered as pdf files to the requesters.

Microbiology laboratory data were imported and analyzed by the use of MOMO (Monitoring of Microorganisms, Medexter Healthcare, Vienna, Austria) [9], a microbiology analytics tool for generating analyses of pathogens, spectra, and antimicrobial resistances from routine microbiology results. In addition, MOMO provides immediate answers to questions related to microbiology results for single patients. This feature is now routinely employed by clinicians at their offices and at the bedside. Figure 1 shows the principal architecture of MOMO. MOMO automatically checks incoming textual identifiers (e.g., specimen, detection method, microbes, antibiotics) for compatibility with existing thesaurus entries, and provides different analysis options. It employs four thesaurus categories: requester/department, specimen type, microbiology, and antibiotics/antimycotics.

Thesaurus Management

MOMO uses software elements which check all incoming entities against the thesaurus they belong to. Importantly, each thesaurus may identify entities either by code or by text depending on its configuration. Due to the structure of MOLIS, MOMO's thesauri for requester/department, specimen type, and antibiotics/antimycotics are based on the entities' respective codes. On the other hand, MOMO's thesaurus for microbiology identifies all entities by their texts because the respective codes in MOLIS were frequently inconclusive or missing.

Regarding the microbiology thesaurus, each of its entries (concepts) consists of an internally generated number and the corresponding textual label/name. Externally provided codes can be attached. MOMO's thesaurus management permits the definition of synonyms for each concept or the creation of new concepts (see Figure 2). The concepts may be organized hierarchically into superordinate and subordinate concepts across several levels (e.g., family, genus, species of bacteria). Concepts may possess more than one parent element. *Staphylococcus aureus*, for instance, is a species of the genus *Staphylococcus* within the family of Staphylococcaceae (parent 1); under the distinction of Gram staining, *Staphylococcus aureus* possesses

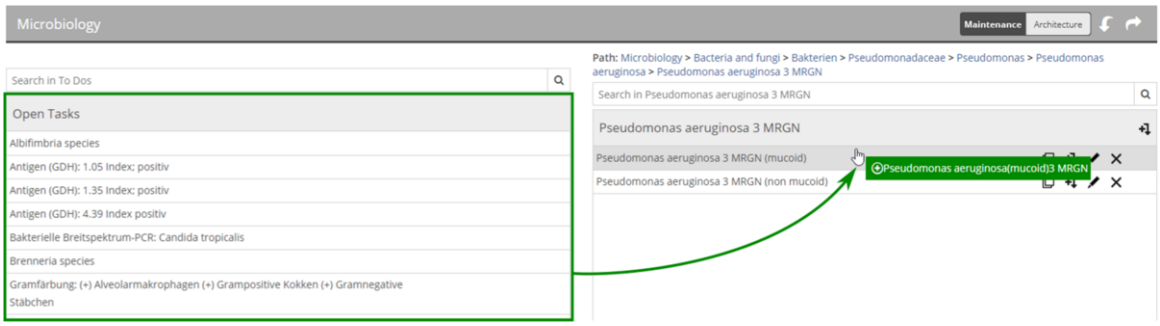

Figure 2 – MOMO thesaurus management – Within MOMO, concepts that cannot be automatically allocated to existing concepts are collected as "open tasks". During thesaurus maintenance these concepts become distinct concepts within the thesaurus hierarchy or will be allocated to an existing concept as a new synonym.

a Gram stain retaining cell wall and is therefore attributed to the Gram-positive group of bacteria (parent 2).

A particular advantage of identifying microbiological results by their texts instead of their codes is that they may be incorporated into the thesaurus immediately, although they may not yet have an officially defined and assigned code [3].

Thus, any microbiology result will be allocated to the respective concept in the thesaurus, or a new concept will be introduced, which is from then on available in clinical as well as secondary query results. Concepts representing misspellings, orthographic variants, or completely new elements must initially be allocated manually; thereafter these texts will be found automatically.

Results

In the investigation period from July 4, 2013 (the day on which the data transfer from MOLIS to MOMO commenced) to February 16, 2018, altogether 773,309 laboratory results were available for analysis. As shown in Table 1, the yearly yield ranged from 154,079 to 209,290, indicating a rising trend. The 2013 and 2018 batches of results were truncated by the respective start and ending dates of the study period.

Table 1 – Imported and analyzed microbiology results by year

Year	Number
2013	25,645
2014	157,627
2015	154,079
2016	196,203
2017	209,290
2018	30,465
Total	773,309

Fractions of Laboratory Results Containing Elements Unknown to the Thesaurus

We started counting the elements unknown to the microbiology thesaurus after the study period had started and can only give an estimate of their total numbers (Table 2).

Roughly 2.6% of all results contained unknown elements and would therefore have been ignored in MOMO analyses based on correctly sorted concepts alone. The types of incompatibility with existing thesaurus concepts were manifold and ranged from typos to completely new elements. Two-thirds required allocation to new sub-concepts under existing concepts. One third of the required actions were the allocation of synonyms or textual variants.

Table 2 – Required thesaurus adaptations after import of 773,309 microbiology results

Type of incompatibility	N*	Fraction
Entries requiring manual allocation	20,000	2.6% of total number
Thereof		
– New species	100	0.5% of N
– New synonym or textual variants	6,900	34.5% of N
– Allocation of new sub-concepts under already existing concepts (especially for serology and microscopy)	13,000	65% of N

* estimated

A prospectively conducted analysis of unassigned microbiology concepts gave a deeper view of the number of concepts allocated to one of the categories within the microbiology thesaurus (Table 3). This analysis includes 89,973 microbiology results from October 30, 2018 to March 12, 2019. Of these, 1,663 (approximately 1.9%) concepts had to be assigned manually. As presented in Table 3, the majority fell into the categories culture, serology, and microscopy.

Table 3 – Current figures for 1,663 manually allocated microbiology concepts between October 30, 2018 and March 12, 2019

Category	N	Fraction
Culture	311	18.7%
PCR	130	7.8%
Serology	525	31,6%
Toxins	72	4.3%
Microscopy	422	25.4%
Miscellaneous	203	12.2%
Total	1,663	100.0%

Effect of the Intervention

On completion of the allocation of unknown elements, 100% of the microbiological results in the trial period became accessible for MOMO analyses.

Manpower Expenditure

In the current use of MOMO, two to three hours per week are required to keep up with changes in laboratory as well as clinical routine. We call this the "man-machine terminology interface". The work is done by persons who are familiar with the MOMO thesaurus as well as clinical and microbiological terms. Top-level clinical or microbiological experts need to be contacted only on rare occasions.

Discussion

Free text entries are known sources of typing errors, misspellings, and unwanted inaccuracies. For this reason, they are "banned" from many modern clinical IT applications, and health care workers are familiar with mandatory coded entries. However, textual descriptions can be more meaningful and may fit individual characteristics better. User-friendly "colloquial" terminologies [2, 4] are indispensable in attracting the use of IT tools by medical professionals.

In the microbiology LIS at the VGH, we were confronted with user demands for free-text elements, which in the course of years of use had become significantly inaccurate. An extraordinary example was a two-digit code connected with 83 – mostly unrelated – entities in the thesaurus. Discrepancies between codes and text were one of the reasons, which had led us to the decision that MOMO should access text rather than codes in the analysis of microbiology terms.

As it turned out, the ambiguity of terms was not an immediate problem for the clinical recipients of microbiology reports, but rather for users relying on MOMO query results, as they would have had to manually select all hidden variations. Hence, the quota of coding incompatibilities observed in our study was more or less irrelevant from the immediate clinical standpoint, but a crucial factor in creating analyses for infection surveillance or outbreak management, where missing results are unacceptable.

As described by de Quirós et al. [2], we had to deal with different acronyms and synonyms for the same clinical finding. For the generation of concise query results, it was necessary to allocate those different terms to the same concept. In contrast to [2], however, we did not have to provide standardized codes based on our thesaurus. Nevertheless, it would be possible to incorporate codes and terms of official terminologies (e.g., SNOMED) as synonyms for those terms that are now available in the microbiology thesaurus.

For this study, we decided not to create a thesaurus based primarily on a so-called reference terminology (e.g., SNOMED as proposed by Rosenbloom et al. [4]) which provides the users with a set of terms as complete as possible. On the contrary, we included only those terms in the thesaurus for which at least one microbiological result was available. Thus, users implicitly know what they may query and do not have to create "test" queries to see whether results are available for certain entries of the thesaurus.

Compliant with, and supplementary to, FAIR principles (code sets are required to be Findable, Accessible, Interoperable, and Reusable) [10], the observations of our study trigger the following microbiology-specific discussion:

Why admit free text entries in lab requests or in (microbiological) lab reports at all?

1. In clinical reporting, we need procedures to deal with inevitable and unanticipated advents of new communication and knowledge elements, which are not yet included in the ordering schemata of the system. This concerns input (emerging clinical demands and laboratory methods) as well as output (new report details or messages). Even brand-new knowledge must be reported precisely and on time, irrespective of the status of the reporting system. This not only applies to microbiology reports for the ordering clinician (which may contain free text information or even handwritten information), but equally to IT analyses and meta-analyses built on microbiology reports, especially if they serve (hospital) epidemiology and outbreak management.
2. Therapeutic imperative: Lab reports must be released a.s.a.p. to facilitate the earliest possible start of appropriate therapy (or adaptation of current therapy) – ideally in a matter of hours. In contrast, as stated by [3], the construction of clinical code sets is usually a time-consuming activity. Thus, the required codes may not be available at the time they are requested.
3. In the man-machine terminology interface, concepts based on free text are organized in a structured manner. This makes them accessible to automated analysis without additional coding.

Following this, the second question pertaining to the analyses of content in microbiology reports is whether they should be based on codes or on free text entries. An analysis based on codes is a plain approach to fully automated analyses; non-coded entries are more difficult to handle and need not be excluded from automated analysis. In free-text-based analyses, each data entry is available for analysis, but the procedure may be time consuming because it requires the scrutiny of hitherto unknown entries.

Neither of the two approaches meets all demands. And, if we agree that non-coded entries have a place in clinical reporting, we ought to provide IT supplements which capture this type of "wild characters" as well.

Our results support two basic recommendations:

1. Optimal and timely thesaurus maintenance is an indispensable provision for all fully automated and rapid IT analyses, the results of which must be trustworthy in terms of their conciseness and correctness.
2. IT systems which draw on codes when appropriate and on free text when appropriate – as MOMO does – provide comfort, speed, and conciseness.

As a result, in thesaurus management high priority is given to capturing synonyms, syntactic deviations and misspellings, which in our analyses were predominant causes of missed entities or misinterpreted reports.

Of similar importance is the ability to capture clinical and microbiological free text comments in laboratory reports which provide special knowledge aimed at new/improved diagnosis and therapy, and which address epidemiology. The ability of a system to communicate such high-level information supports the proficiency/professionality of the daily dialog between laboratory, clinicians and hospital epidemiologists.

Finally, this type of IT tool may serve as an internal knowledge engine for the clinical microbiology laboratory striving for continuous knowledge acquisition and its provision in routine microbiological work.

This topic also encompasses the notorious question: can analyses of routine clinical reporting be accepted for research purposes, especially in epidemiology and public health matters?

This question was discussed extensively in the process of establishing international benchmarking networks (e.g., HELICS [11], IPSE [12], and EARSS [13], which were forerunners of the present European Centre for Disease Prevention and Control networks HAI-net [14] and EARS-net [15]). In the end, it was agreed that routine laboratory data are indispensable for this purpose, despite the fact that they may be generated under missing or unknown scientific standards. There is no other way of obtaining the required information, because impeccable scientific studies of appropriate size and duration are not feasible. In our study, we present a new focus as well as a software solution to bring analysis data from routine clinical reporting closer to the desired degree of conciseness, reliability, and relevance.

Conclusions

In our opinion, clinical IT solutions must focus on a good balance between full automation and man-machine interactivity for successful clinician-laboratory dialog, which in turn supports patient care and infection control.

Given this balance, we may expect considerable progress from such IT solutions in microbiology-related "Good Clinical Practice" as well as in infection prevention and control. Other areas that will benefit from such progress are research in clinical microbiology, healthcare-associated infection prevention and control, epidemiology, and public health!

"When used properly, informatics tools can help the clinical microbiology laboratory to do more with less while improving the quality of patient and public health care" [1].

Acknowledgments

The authors are indebted to Wolfgang Barousch for his fruitful help and discussions.

References

[1] D.D. Rhoads, V. Sintchenko, C.A. Rauch, and L. Pantanowitz, Clinical Microbiology Informatics, *Clin Microbiol Rev* **27**(4) (2014), 1025-1047.

[2] F.G.B. de Quirós, C. Otero, D. Luna, Terminology Services: Standard Terminologies to Control Health Vocabulary, *Yearb Med Inform* **27**(1) (2018), 227-233.

[3] R. Williams, B. Brown, E. Kontopantelis, T. van Staa, N. Peek, Term Sets: A Transparent and Reproducible Representation of Clinical Code Sets, *PloS On*e **14**(2) (2019) :e0212291.

[4] S.T. Rosenbloom, R.A. Miller, K.B. Johnson, P.L. Elkin, S.H. Brown, Interface Terminologies: Facilitating Direct Entry of Clinical Data into Electronic Health Record Systems, *J Am Med Inform Assoc* **13**(3) (2006), 277-288

[5] 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, *J Hosp Infect* **84**(2) (2013), 106-119.

[6] W. Koller, K.-P. Adlassnig, A. Rappelsberger, A. Blacky, Plea for Use of Intelligent Information and Communication Technologies in Infection Surveillance and Benchmarking by Healthcare Institutions, in M. Bienkiewicz, C. Verdier, G. Plantier, T. Schultz, A. Fred, H. Gamboa (Eds.) *Proceedings of HEALTHINF 2014*, Scitepress, Portugal, (2014) 399-404.

[7] W. Koller, J. de Bruin, W. Barousch, B. Willinger, K.-P. Adlassnig, Man-Machine Dialog: How to Optimize Results of Intelligent IT Tools in Infection Surveillance and in Clinical Decision Support, *Int J Infect Control* **14**(Supp 1: Abstracts of IFIC 2018, 25–27 April 2018, Krakow, Poland) (2018), 10.

[8] CompuGroup Medical (CGM) SE, *MOLIS – Laboratory Information System,* 2017. Available at: https://www.cgm.com/media/cgm_de/documents/labor_1/2017-01-24_MOLIS_Brosch.pdf, last access: 21 October 2018.

[9] Medexter Healthcare GmbH, *Microbiology Analytics with MOMO – Monitoring and Reporting Pathogens & AMR,* 2018. Available at: https://www.medexter.com/products-and-services/clinical-solutions/microbiology-and-amr, last access: 28 March 2019.

[10] M.D. Wilkinson, M. Dumontier, U. Aalbersberg I.J., G. Appleton, M. Axton, A. Baak, et al., The FAIR Guiding. Principles for Scientific Data Management and Stewardship, *Sci Data* **3** (2016) :160018.

[11] HELICS – Hospital in Europe Link for Infection Control through Surveillance, *Surveillance of Nosocomial Infections in Intensive Care Units. Protocol Version 6.1 (Based on Version 5.0 including technical amendments) September 2004*, 2004. Available at: http://www.sicsag.scot.nhs.uk/hai/helics_protocol.pdf, last access: 28 March 2019.

[12] European Centre for Disease Prevention and Control, *Improving Patient Safety in Europe. Technical Implementation Report 2005-2008. November 2008, Vol. I*, 2008. Available at: https://ecdc.europa.eu/sites/portal/files/media/en/healthtopics/Healthcare-associated_infections/HAI-Net/Documents/healthcare-associated-infections-IPSE-Technical-Report.pdf, last access: 28 March 2019.

[13] National Institute for Public Health and the Environment, *AMR Surveillance Infrastructure*, 2015. Available at: https://www.rivm.nl/en/Topics/W/WHO_Collaborating_Centre_Antimicrobial_Resistance_Epidemiology_and_Surveillance/Fields_of_expertise/AMR_surveillance_infrastructure, last access: 28 March 2019.

[14] European Centre for Disease Prevention and Control, *About the Network*, 2018. Available at: https://ecdc.europa.eu/en/about-us/networks/disease-networks-and-laboratory-networks/hai-net-about, last access: 28 March 2019.

[15] European Centre for Disease Prevention and Control, *European Antimicrobial Resistance Surveillance Network (EARS-Net),* 2018. Available at: https://ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/ears-net, last access: 28 March 2019.

Address for Correspondence

Univ.-Prof. Dr. Walter Koller, Medical University of Vienna; Mail c/o Martinstrasse 26, 1180 Vienna, Austria; Tel.: +43-676-5008106; E-Mail: walter.koller@meduniwien.ac.at.