A Multi-User Terminology Mapping Toolbox

Sebastian MATE [a,1], Susanne A. SEUCHTER [a],
Katharina EHRENBERG [a], Noemi DEPPENWIESE [b], Jakob ZIERK [a,c],
Hans-Ulrich PROKOSCH [a,b], Detlef KRASKA [a], and Lorenz A. KAPSNER [a]

[a] *Medical Center for Information and Communication Technology,*
Universitätsklinikum Erlangen, Erlangen, Germany
[b] *Chair of Medical Informatics, Friedrich-Alexander-Universität*
Erlangen-Nürnberg (FAU), Erlangen, Germany
[c] *Department of Pediatrics and Adolescent Medicine,*
Universitätsklinikum Erlangen, Erlangen, Germany

Abstract. Semantic interoperability is a major challenge in multi-center data sharing projects, a challenge that the German Initiative for Medical Informatics is taking up. With respect to laboratory data, enriching site-specific tests and measurements with LOINC codes appears to be a crucial step in supporting cross-institutional research. However, this effort is very time-consuming, as it requires expert knowledge of local site specifics. To ease this process, we developed a generic manual collaborative terminology mapping tool, the *MIRACUM Mapper*. It allows the creation of arbitrary mapping workflows involving different user roles. A mapping workflow with two user roles has been implemented at University Hospital Erlangen to support the local LOINC mapping. Additionally, the *MIRACUM LabVisualizeR* provides summary statistics and visualizations of analyte data. We developed a toolbox that facilitates the collaborative creation of mappings and streamlines the review as well as the validation process. The two tools are available under an open source license.

Keywords. Secondary Use, Interoperability, Laboratory, LOINC, Terminology

1. Introduction

With the introduction of electronic health records, medical information has become widely available in the form of electronic data. There is great interest in reusing this data for research, which is usually referred to as "secondary use." In recent years, many initiatives and research networks have emerged for sharing data for secondary use, such as *PCORnet* [1], *OHDSI* [2], or *SHRINE* [3]. In Germany, the *Medical Informatics Initiative* (MII) was initiated with four consortia, which are funded since 2018 [4].

As mentioned by several authors (e. g. [5]), semantic interoperability is a major problem for such research networks. To address this challenge within the MII, a working group on interoperability has been established. One of its milestones includes the definition of a MII core data set with seven basic modules, of which laboratory data are part. Within the MII, the national steering committee has agreed on a list of the 300 most commonly used laboratory tests, primarily mapped to the LOINC standard [6]. Consequently, linking site-specific tests and measures to LOINC is an important task step to

---

[1]   Corresponding Author. Postal Address: Universitätsklinikum Erlangen, Medizinisches Zentrum für Informations- und Kommunikationstechnik, Krankenhausstraße 12, 91054 Erlangen, Germany; E-Mail: Sebastian.Mate@uk-erlangen.de

improve interoperability in the laboratory field, a task to be pursued by so-called *Data Integration Centers* (DICs). Established at German university hospitals, these DICs are the organizational units to process and provide medical data for research, such as diagnoses, procedures, medication, laboratory data and billing-related data [4]. However, as mentioned by Fiebeck et al. [7], the manual mapping of local terms to the LOINC standard is a very time-consuming process, which also requires significant expert knowledge of site-specific local laboratory procedures.

At University Hospital Erlangen (UHE), which is part of the MII-funded MIRA-CUM consortium [8], a semi-automatic LOINC mapping with RELMA V.5 has been performed by Zunner et al. [9] in the past. More than 2,000 local laboratory terms were mapped to the respective LOINC codes and the associated evaluation study [9] has shown good results. At that time, however, no routine process could be established in cooperation with the laboratories at UHE in which the current, but also future or modified local laboratory codes could have been mapped to LOINC under strict quality control. Routine laboratory personnel usually do not have the resources (in particular time and software tooling) to carry out such a mapping project for thousands of codes.

The task of creating mappings in general is a recurring challenge. Regarding LOINC, there are tools such as RELMA [10], however, to the best of our knowledge, there is no tool yet that allows multiple users to work in parallel on a mapping project. To tackle this problem, we implemented a generic collaborative terminology mapping application. It was then used at UHE by a specifically assigned mapping expert who then mapped all local laboratory codes to LOINC. These were subsequently validated by the laboratory personnel. In this paper, we describe the design and the implementation of this multi-user application and report on its current use at UHE.

2. Methods

The design of the mapping tool considered the following requirements:

- It should be executable on the PC systems commonly found in German hospitals.
- It should support central user authentication for keeping user management simple.
- It should support the versioning of mappings and an audit trail in which all mapping steps and mapping validations are logged.
- It should enable multiple users to work asynchronously on different mappings.
- It should primarily support the laboratory mapping task, but should also be generic and configurable enough to be adapted for future harmonization tasks in other domains. A generic approach to support arbitrary mapping workflows, including support for different types of users, would be desirable.

We implemented the mapping tool, *MIRACUM Mapper*, as a multi-user C# Windows application. It uses the *lightweight directory access protocol* (LDAP) to authenticate its users and connects to a central PostgreSQL database, where it stores the versioned mappings from the local laboratory system's analysis codes to the LOINC terminology. The application can be configured for arbitrary mapping workflows and different user roles via an internal, configurable state machine. Figure 1 depicts the state transitions for the two user roles "expert" and "laboratory", as implemented in the LOINC mapping use case at UHE. It comprises six states, as depicted with the boxes. The state transitions for

a certain user group are modeled via transition lists. Transitions are defined using tuples in the format *m-n*, where *m* denotes the start state and *n* the end state.
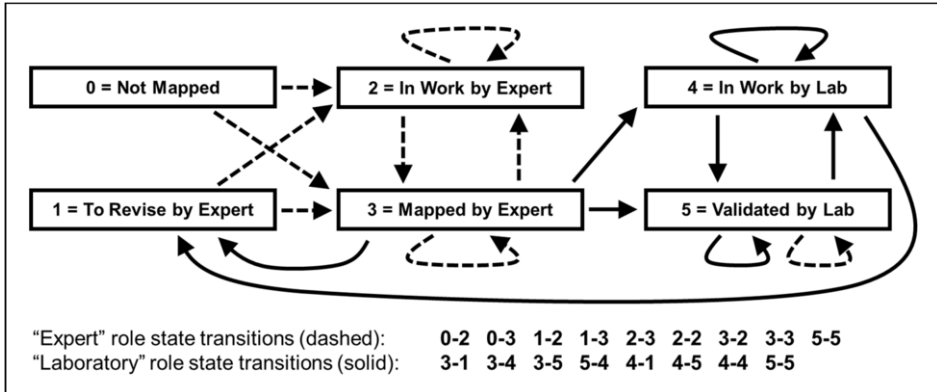

| "Expert" role state transitions (dashed): | 0-2 0-3 1-2 1-3 2-3 2-2 3-2 3-3 5-5 |
| "Laboratory" role state transitions (solid): | 3-1 3-4 3-5 5-4 4-1 4-5 4-4 5-5 |

Figure 1. State transitions used in the LOINC mapping use case at UHE.

When preparing the system for the UHE LOINC mapping project, we assigned all unmapped local codes to state 0. The previously made mappings (from [9]) were imported as state 1 mapping in order to be revised again. The entries in state 0 and 1 were then to be processed by the mapping expert with the goal of achieving state 3. Mappings in state 3 were then picked up by the laboratory users, with the ultimate goal of achieving state 5. Notice that in state 4 the laboratory users are also able to "send" mappings back to state 1, thereby asking the expert users to revise them again. The loops (2-2, 3-3, 4-4 and 5-5) additionally allow users to edit the documentation without changing the state.

3. Results

A screenshot of MIRACUM Mapper is shown in Figure 2. The program displays the "source" terminology on the left side, which in the LOINC mapping use case is a list of local laboratory codes. This list also serves as a "to do" list: depending on the filter settings in the lower left corner, the program displays different local codes. In the UHE workflow configuration (as shown in Figure 1), the mapping experts by default see the local codes for which there are no LOINC mappings or where mappings are to be revised (states 0 and 1), whereas the laboratory users see the mappings "finished" by the experts (state 3). Local codes with recently edited mappings are always shown on top of the list to indicate that the mapping has been updated by one of the other users.

Mappings to the "target" LOINC terminology must initially be researched outside the tool. Consequently, the mapping task relies heavily on external resources, such as the LOINC website. If a suitable target code has been found, a new mapping can be defined in the center of the window. The LOINC code is then entered into the "Target Code" field in the upper right corner. For quality control, however, the program performs an automatic search in the target terminology to ensure that the code entered is valid. In addition, the "Search" button can be used to open a browser with a preconfigured URL to view further details for a given target code. In the LOINC setting, this opens the website for the target code on the LOINC website (in the example, https://loinc.org/2075-0/).
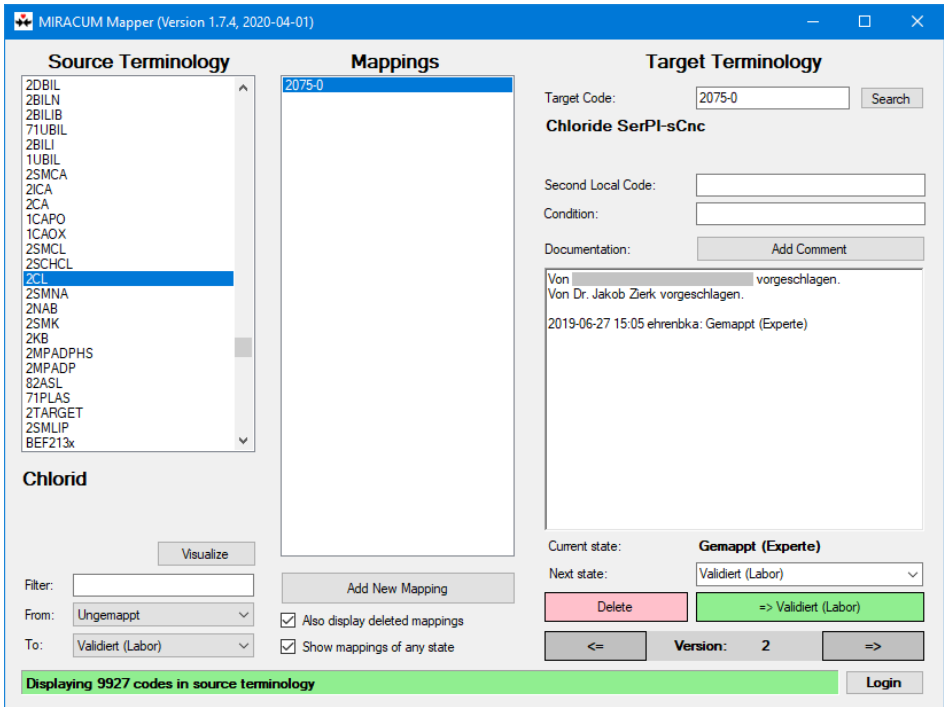
Figure 2. User interface of MIRACUM Mapper.

The documentation field on the right hand side can be used for making notes during the mapping process, which are visible to all users. Also, automatically generated entries with timestamps further keep track of the editing process. The tool automatically detects whenever an entered target code was already used in a previous or deleted mapping, even if the program is set to not display deleted mappings (notice the checkbox in the lower part of the window). In both cases, it recalls the previous documentation field and thereby restores the previously deleted mapping. When saving the mapping, the internal state machine of the tool transfers the mapping to the next state.

After the initial rollout, we realized that it would be helpful for users to be able to review additional information about local codes. As this is a requirement specific to the laboratory task, we developed another application called *MIRACUM LabVisualizeR* (see Figure 3). It aims to provide quick access to additional information available in the original laboratory system, including metadata, anonymized sample patient data, but also computed statistics. It is implemented in the "R" programming language [11] using the "shiny" package [12]. It can be opened directly from the MIRACUM Mapper's user interface by pressing the "Visualize" button. It displays histograms, violin plots, and summaries for the currently selected laboratory code. The presented results are derived from the most recent 100,000 values of the respective data element retrieved from the UHE data warehouse. Outliers can be removed by ticking a checkbox or by adjusting the value range using a slider. Similarly, an underlying gender cohort can be specified using radio buttons. A second slider is available for constraining the range of the cohort's age.

Both programs were used by a LOINC expert at UHE for the mapping of 9,970 local laboratory terms to LOINC. This process took about six months with the expert working full-time. The validation of the mappings by the UHE's laboratories is an ongoing process. As of July 16th, 2020, 5,721 mappings were validated by currently 22 members from 11 different UHE laboratories. Currently other MIRACUM-affiliated university hospitals in Germany consider using the tool for updating their LOINC mappings.
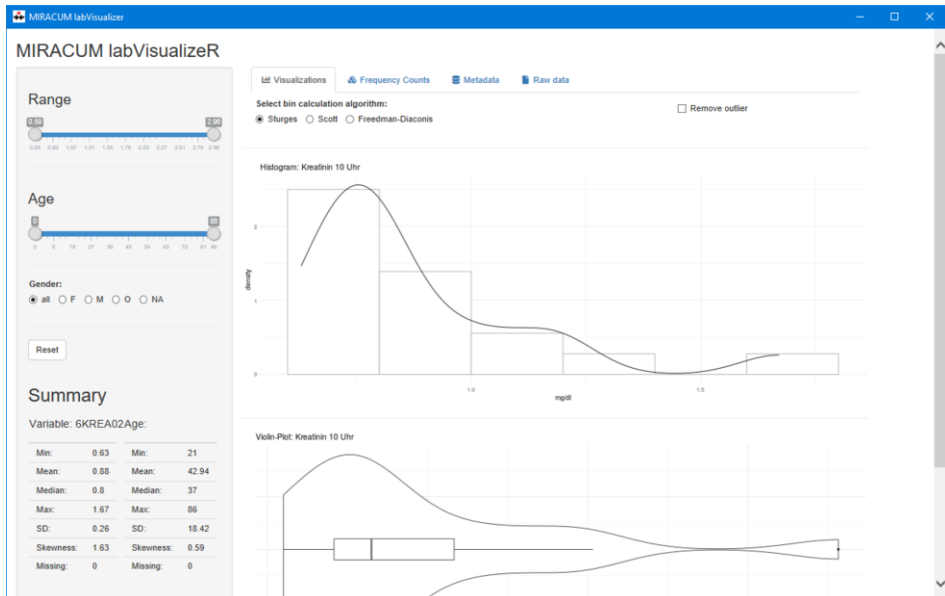

Figure 3. User interface of MIRACUM LabVisualizeR.

4. Discussion and Outlook

We have presented a toolbox for the manual mapping of terminologies. It provides a single consistent user interface and streamlines the entire process of creating, checking, and validating mappings. Although we have not yet conducted a formal user evaluation, initial feedback suggests that the application is perceived well. We are currently comparing the new mappings with the former ones [9] in detail and plan to report on our findings, including details on user feedback, in the near future. The result of the UHE LOINC mapping will support many secondary use projects in the future. In addition, it could act as a gold standard for evaluating future research on semi-automatic mapping approaches.

We recently developed a semi-automatic mapping approach within the ADOPT BBMRI-ERIC biobank networking project [13] and considered using it for the LOINC mapping. We updated the matching component with multi-threading capabilities and an improved matching algorithm to handle the very large LOINC terminology, but initial matching results were not satisfactory. We assume that this is due to the high heterogeneity of data elements present in LOINC. We also considered using RELMA [10], which offers an automatic suggestion of mappings, but unfortunately does not allow multi-user

workflows as mentioned in our requirements. We suspected that most of the work consists of researching content, but not in selecting the "best" match. Our current experience supports this: the time required to create mappings varied depending on its "difficulty" level. Apparent terms, such as "glucose in serum", were mapped in less than a minute, whereas others (e.g., different types of antibodies) took longer, because they required literature research or communication with laboratory experts or physicians. It is not clear how these tasks can be further supported via such semi-automatic mapping extensions.

Both tools are available under the GPL3 open source license and can be downloaded from https://gitlab.miracum.org/matesn/miracum-mapper and https://gitlab.miracum.org/miracum/labor/labVisualizeR.

Declarations

References

[1]   F.S. Collins, K.L. Hudson, J.P. Briggs, and M.S. Lauer, PCORnet: Turning a Dream Into Reality, *J. Am. Med. Inform. Assoc.* **21** (2014) 576–577. doi:10.1136/amiajnl-2014-002864.

[2]   G. Hripcsak, J.D. Duke, N.H. Shah, C.G. Reich, and V. Huser, Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers, *Stud. Health Technol. Inform.* (2015). doi:10.3233/978-1-61499-564-7-574.

[3]   A.J. McMurry, S.N. Murphy, D. MacFadden, G.M. Weber, W.W. Simons, J. Orechia, J. Bickel, N. Wattanasin, C. Gilbert, P. Trevvett, S.E. Churchill, and I.S. Kohane, SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies, *PloS One.* **8** (2013) e55811. doi:10.1371/journal.pone.0055811.

[4]   S.C. Semler, F. Wissing, and R. Heyder, German Medical Informatics Initiative, *Methods Inf. Med.* **57** (2018) e50–e56. doi:10.3414/ME18-03-0003.

[5]   D. Teodoro, R. Choquet, E. Pasche, J. GOBEILL, C. Daniel, P. Ruch, and C. Lovis, Biomedical Data Management: a Proposal Framework, *Stud. Health Technol. Inform.* **150** (2009) 175–179.

[6]   A.W. Forrey, C.J. McDonald, G. DeMoor, et al., Logical Observation Identifier Names and Codes (LOINC) Database: A Public Use Set of Codes and Names for Electronic Reporting of Clinical Laboratory Test Results., *Clin. Chem.* **42** (1996) 81–90.

[7]   J. Fiebeck, M. Gietzelt, S. Ballout, M. Christmann, M. Fradziak, H. Laser, J. Ruppel, N. Schönfeld, S. Teppner, and S. Gerbel, Implementing LOINC - Current Status and Ongoing Work at a Medical University, *Stud. Health Technol. Inform.* **267** (2019) 59–65. doi:10.3233/SHTI190806.

[8]   H.-U. Prokosch, T. Acker, J. Bernarding, H. Binder, M. Boeker, M. Boerries, P. Daumke, T. Ganslandt, J. Hesser, G. Höning, M. Neumaier, et al., MIRACUM: Medical Informatics in Research and Care in University Medicine, *Methods Inf. Med.* **57** (2018) e82–e91. doi:10.3414/ME17-02-0025.

[9]   C. Zunner, T. Bürkle, H.-U. Prokosch, and T. Ganslandt, Mapping Local Laboratory Interface Terms to LOINC at a German University Hospital Using RELMA V.5: A Semi-automated Approach, *J. Am. Med. Inform. Assoc.* **20** (2013) 293–297. doi:10.1136/amiajnl-2012-001063.

[10]   D.J. Vreeman, J. Hook, and B.E. Dixon, Learning from the Crowd While Mapping to LOINC., *J. Am. Med. Inform. Assoc. JAMIA.* **22** (2015) 1205–1211. doi:10.1093/jamia/ocv098.

[11]   R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2019.

[12]　W. Chang, J. Cheng, J.J. Allaire, Y. Xie, and J. McPherson, Shiny: Web Application Framework for R, 2019.

[13]　S. Mate, M. Kampf, W. Rödle, S. Kraus, R. Proynova, K. Silander, L. Ebert, M. Lablans, C. Schüttler, C. Knell, N. Eklund, M. Hummel, P. Holub, and H.-U. Prokosch, Pan-European Data Harmonization for Biobanks in ADOPT BBMRI-ERIC, *Appl. Clin. Inform.* **10** (2019) 679–692. doi:10.1055/s-0039-1695793.