



UNIVERSITÄTS
KLINIKUM
HEIDELBERG

Publication Series:
Heidelberg Spring Symposium Medical
Informatics

IMI INSTITUT FÜR
MEDIZINISCHE
INFORMATIK



Petra Knaup, Martin Dugas, Matthias Ganzinger (eds.)

1. Heidelberger Frühjahrssymposium Medizinische Informatik

17.05.2023

Abstractband

This publication is the first volume of a series of publications by the Institute of Medical Informatics, Heidelberg University Hospital (Germany). The series contains the abstract volumes of the annual Heidelberg Spring Symposium Medical Informatics, which has been taking place since 2023.

Acknowledgements

The editors would like to thank all the staff of the Institute who contributed to the success of the symposium with their commitment. Organizing a symposium is always a team effort, but we would especially like to thank Karin Schmid and Veronica Wilhelm for their excellent organizational support.

Cite as:

Knaup P, Dugas M, Ganzinger M (2024): 1. Heidelberger Frühjahrssymposium Medizinische Informatik. Heidelberg University Library. <https://doi.org/10.11588/heidok.00034809>

Contact:

Institute of Medical Informatics (Teamassistenz-Postfach.IMI@med.uni-heidelberg.de)

Published by Heidelberg University Library, 2024

The electronic version of this work is permanently available on:
<https://archiv.ub.uni-heidelberg.de/volltextserver/>

doi: <https://doi.org/10.11588/heidok.00034809>

Text © 2024, Petra Knaup

Inhaltsverzeichnis

Hauptvorträge

H. Abu Attieh et al. (Berlin): PROMs und DiGAs: Chancen und Herausforderungen	6
D. Krefting et al. (Göttingen): NUKLEUS - A Technical and Organisational Research Infrastructure to Support Timely, High Quality and FAIR Clinical Studies	7
M. Ganzinger et al. (Heidelberg): Föderierte Datenerfassung für die klinische Forschung	9
P. Richter-Pechanskii et al. (Heidelberg): Section Classification in German Discharge Letters using Few-Shot Learning and Prompting	11

Impulsvorträge

Th. Deutsch et al. (Heidelberg): ENABLE - innovative care concepts in the treatment of breast cancer patients	14
E. M. Hartmann et al. (Dortmund): Klinisches Informationssystem OPICAD zur Versorgungsoptimierung schwerstkranker Kinder aus aller Welt	15
J. Bittmann et al. (Heidelberg): Entwicklung einer neuen Methodik zur automatisierten Bestimmung der Akzeptanz von Medikations-Alerts mittels Ereignisanalyse	17
L. Meier (Cambridge, UK): Medical Informatics meets medical ethics	19
M. Botelho et al. (Tübingen): Secure Distributed Medical Analytics for German Healthcare institutions using the Personal Health Train (PHT-meDIC)	20
C. Bruns et al. (Magdeburg): Using DICOM headers in FHIR Database for enrichment of the Medical Informatics Initiativ's core data set	21
A. Pfob et al. (Heidelberg): IMI-EDC Pilot Implementation at the Department of Obstetrics and Gynecology Heidelberg	23
T. Ohlsen et al. (Lübeck): Webanwendung zur Unterstützung einer SNOMED CT-basierten Postkoordination	24
D. Hüske-Kraus (Böblingen): Ontologiebasierte Erzeugung klinischer Dokumente	25

Poster

M. Aigner et al. (Neu-Ulm): Mobile Apps for COVID 19: First Results of a Systematic Review of Reviews	28
J. Bickenbach et al. (Aachen): Use Case ASIC	30
C. Bönisch et al. (Göttingen): Make Metadata relational again: Entwicklung einer Metadatenstruktur innerhalb des UMG-MeDIC	31
S. Hegde et al. (Heidelberg): Interactive Feasibility and Data Explorer for Clinical Data (FEDEX) ..	33
Y. Hollenbenders et al. (Heilbronn): Robustness of Electroencephalography-Biomarkers for Major Depressive Disorder – an Exemplary Study with Alpha Bandpower	34
M. Karthan et al. (Neu-Ulm): Developing a Mobile Serious Game Platform to Improve Orofacial Myofunctional Therapy for Children	36
D. Molinnus et al. (Aachen): Entwicklung einer digitalen Infrastruktur für eine intersektorale medizinische Versorgung – der klinische Anwendungsfall PICOS.....	37
E. Thomas et al. (Jena): Entwicklung von ETL-Verfahren für die interoperable Nutzung von Gesundheitsdaten – Grundprinzipien und Lessons learned	38
J. Varghese et al. (Münster): OpenEDC - Für interoperable und dezentrale Forschung	41
R. Verbücheln et al. (Tübingen): Verteilte Analysen zur Untersuchung der aktuellen Dokumentation ausgewählter seltener Erkrankungen zwischen verschiedenen Krankenhäusern mit DataSHIELD	42
S. Welten et al. (Aachen): Function, Aim, and Merit of the Personal Health Train for the Secondary Use of Clinical Data in Research	43



Hauptvorträge

Titel: PROMs und DiGAs: Chancen und Herausforderungen

Autoren: Hammam Abu Attieh, Matthias Rose, Fabian Prasser

Abstract: Die Integration von Patient-Reported Outcome Measures (PROMs) in die klinische Praxis bietet das Potential, die Kommunikation zwischen Arzt und Patient zu verbessern, Behandlungsprozesse zu optimieren und den Nutzen von Behandlungen besser zu verstehen (1). Viele Projekte arbeiten deshalb daran, PROMs mithilfe digitaler Fragebögen und Apps effektiver zu erfassen und zusammen mit klinischen Daten auszuwerten (2). Dazu bedarf es einer skalierbaren Infrastruktur, welche die Vernetzung relevanter Akteure sicherstellt und die PROMs verwaltet. Diese ist grundsätzlich mit der Telematikinfrastruktur (TI) und der darauf aufbauenden elektronischen Patientenakte (ePA) gegeben, die auch Daten aus sogenannten Digitalen Gesundheitsanwendungen (DiGAs) integrieren kann. Für die Nutzung der TI bzw. der ePA für DiGAs, welche die Erfassung von PROMs als Hauptfunktion implementieren, gibt es jedoch technische und rechtliche Hürden.

Zwar wurde eine allgemeine Grundlage für den Datenexport von DiGAs in die ePA mit Art. 351 Abs. 2 SGB V geschaffen. Gemeint sind hier jedoch lediglich DiGAs, die gem. Art. 33a SGB V in das sogenannte DiGA-Verzeichnis des Bundesinstituts für Arzneimittel und Medizinprodukte (BfArM) aufgenommen wurden. Diese müssen neben Anforderungen an die Sicherheit, Funktionstauglichkeit und Qualität nach Art. 3 bis 6 der Digitale-Gesundheitsanwendungen-Verordnung (DiGAV) einen „positiven Versorgungseffekt“ nachweisen (3). Dies erschwert die Entwicklung von DiGAs, welche ausschließlich auf die Erfassung von PROMs ausgelegt sind, da diese den entsprechenden Anforderungen nur schwer gerecht werden können. Zur Sicherstellung der Interoperabilität wurde die Kassenärztliche Bundesvereinigung (KBV) gem. Art. 355 SGB V mit dem Entwurf von medizinischen Informationsobjekten (MIO) beauftragt. Eine herstellerunabhängige MIO-Spezifikation (MIO DiGA Toolkit) wurde in Abstimmung mit den Herstellern der im Verzeichnis des BfArM gelisteten DiGAs entwickelt. Das Toolkit basiert auf generischen Modulen, die sowohl persönliche Angaben und Metadaten, als auch optionale Informationen entsprechend der Anforderungen der jeweiligen DiGA abbilden können. In Bezug auf PROMs existieren beispielsweise Module für Fragebögen oder Patientenberichte. Die Module definieren technische Anforderungen an die Interoperabilität des zu integrierenden Datensatzes für DiGAs i.S.v. Art. 33a SGB V basierend auf HL7 FHIR. Jedoch sind die einzelnen Spezifikationen nicht immer vollständig aufeinander abgestimmt. Darüber hinaus nutzt bisher nur etwa ein Drittel der DiGAs die bestehenden MIOs (Stand: März 2022) (4), da die Umsetzung teilweise sehr komplex ist (5).

Literaturverzeichnis:

1. Stamp E, Clarke G, Wright P, Velikova G, Crossfield SSR, Zucker K, et al. Collection of cancer Patient Reported Outcome Measures (PROMS) to link with primary and secondary electronic care records to understand and improve long term cancer outcomes: A protocol paper. PLoS One. 2022;17(4):e0266804.
2. Stamm T, Bott N, Thwaites R, Mosor E, Andrews MR, Borgdorff J, et al. Building a value-based care infrastructure in Europe: the health outcomes observatory. NEJM Catalyst Innovations in Care Delivery. 2021;2(3).
3. BfArM. Das Fast-Track-Verfahren für digitale Gesundheitsanwendungen (DiGA) nach § 139e SGB V—Ein Leitfaden für Hersteller, Leistungserbringer und Anwender. 2022; S.39.
4. Techniker Krankenkasse. DiGA-Report 2022. 2022; S.27.
5. bvitg e.V. Gematik: Weiterentwickeln, aber richtig! bvitg-Monitor. 2022;(4/22): S.54–61.

NUKLEUS - A Technical and Organisational Research Infrastructure to Support Timely, High Quality and FAIR Clinical Studies

Dagmar KREFTING^{a,b,1}, Gabi ANTON^c, Irina CHAPLINSKAYA-SOBOL^a, Sabine HANSS^a, Wolfgang HOFFMANN^d, Sina M. HOPFF^{e,f}, Monika KRAUS^c, Roberto LORBEER^g, Bettina LORENZ-DEPIEREUX^c, Thomas ILLIG^h, Christian SCHÄFERⁱ, Jens SCHALLER^g, Anne SCHONEBERG^a, Dana STAHL^j, Heike VALENTIN^j, Peter HEUSCHMANN^k and Janne VEHRESCHILD^{e,1}

^a*Dpt. of Medical Informatics, University Medical Center Göttingen, German Center for Cardiovascular Research (DZHK) partner site Göttingen, Germany*

^b*Campus Institute Data Science (CIDAS), Georg-August-University Göttingen,*

^c*Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany*

^d*Institute for Community Medicine, University Medicine Greifswald, Germany*

^e*Faculty of Medicine, University of Cologne, Department I of Internal Medicine, University Hospital Cologne, Germany*

^f*Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Cologne, Germany*

^g*Deutsches Herzzentrum der Charité, Charité – Universitätsmedizin Berlin, Institute of Computer-assisted Cardiovascular Medicine, German Center for Cardiovascular Research (DZHK) partner site Berlin, Germany*

^h*Hannover Unified Biobank, Hannover Medical School, Hannover, Germany*

ⁱ*Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Germany*

^j*Independent Trusted Third Party of the University Medicine Greifswald, Germany*

^k*Institute of Clinical Epidemiology and Biometry, University of Würzburg; Clinical Trial Center, University Hospital Würzburg, Germany*

¹*German Centre for Infection Research (DZIF), partner site Bonn- Cologne, Cologne, Department II for Internal Medicine, Hematology/Oncology, University Hospital Frankfurt, Frankfurt am Main, Germany.*

The Network University Medicine Clinical and Epidemiology Platform (NUKLEUS) has been founded 2022 as a long-term research infrastructure. It aims at supporting researchers to efficiently plan and conduct clinical-epidemiological studies and enable broad reuse of the study data. NUKLEUS combines the methodological core units of the National Pandemic Cohort Network (NAPKON [1]) and the study infrastructure adopted from the German Center for Cardiovascular Research [2], funded within the Network University Medicine (NUM [2]). Currently, NUKLEUS consists of four dedicated data management systems: The clinical data capture system secuTrial® for eCRFs, the laboratory information system CentraXX® for biospecimen metadata and Trialconnect® for DICOM data are connected by the trusted third party managing identifying data, digital consents and system-specific pseudonyms with the open source tools gICS®, E-PIX® and gPAS®. Performance and data quality reports are automatically generated on aggregated data and delivered to lead investigators and participating sites. Fast implementation of a study is supported by coordination of harmonized study and proband documents for ethics approval, transparent cost calculations for using NUKLEUS, and provision of an up-and-running research data infrastructure. Epidemiological consultancy is offered throughout the full study lifetime – from the planning phase to the statistical analysis. The Interaction Core Unit (ICU) coordinates internal and external communication and implements and optimizes the required processes. The Biosample Core Unit (BCU) ensures high-quality biospecimen collection including site audits. The Epidemiology Core Unit is responsible for quality assurance. NUKLEUS is specifically designed to implement and support FAIR data stewardship [5]. Data usage requests are managed through the Proskive application management system, Use and Access committees are organized by the ICU and data provision is carried out by the transfer office – for biospecimen the BCU. Again methodological consultancy is offered for the data request application. To date, more than 100 usage applications have been approved, of which so far 55 received data. About 32.000 reviewed visits from more than 6.500 NAPKON

¹ Corresponding Author: Dagmar Krefting, University Medical Center Göttingen, Germany, dagmar.krefting@med.uni-goettingen.de

participants are ready for reuse. From a total of 83.300 primary biospecimen, 12 sample usage applications received a total of 36.600 sample aliquots.

[1] Schons M, Pilgram L, Reese JP, Stecher M, Anton G, Appel KS, et al. The German National Pandemic Cohort Network (NAPKON): rationale, study design and baseline characteristics. *Eur J Epidemiol* [Internet]. 2022 Jul 29 [cited 2022 Aug 1]; Available from: <https://doi.org/10.1007/s10654-022-00896-z>

[2] Deutsches Zentrum für Herz-Kreislaufforschung e.V. DZHK Heart Bank: DZHK [Internet]. [cited 2023 Jan 31]. Available from: <https://dzhk.de/en/dzhk-heart-bank/>

[3] Heyder R, NUM Coordination Office, Kroemer HK, Wiedmann S, Pley C, Heyer C, et al. Das Netzwerk Universitätsmedizin: Technisch-organisatorische Ansätze für Forschungsdatenplattformen. *Bundesgesundheitsbl* [Internet]. 2023 Jan 23 [cited 2023 Jan 24]; Available from: <https://link.springer.com/10.1007/s00103-022-03649-1>

[4] Institute of Medical Informatics Frankfurt. ProSkive [Internet]. [cited 2023 Jan 31]. Available from: <https://proskive.de/en>

[5] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016 Mar 15;3:160018.

Föderierte Datenerfassung für die klinische Forschung

Matthias Ganzinger, Max Blumenstock, Christian Niklas, Martin Dugas
Institut für Medizinische Informatik, Universitätsklinikum Heidelberg

Elektronische Datenerfassungssysteme (engl. Electronic Data Capture, *EDC*, Systems) werden häufig in der klinischen Forschung sowie in Studien eingesetzt [1]. Diese Systeme haben die Aufgabe, elektronische Formulare für die strukturierte Erfassung von Beobachtungsdaten bereitzustellen und deren Verarbeitung zu unterstützen. Für diesen Zweck steht eine Vielzahl an kommerziellen, aber auch kostenfreien open-source Lösungen zur Verfügung. Dabei werden sowohl Erhebungen unterstützt, die sich auf eine einzelne Einrichtung beziehen, aber auch multizentrische Studien, an denen mehrere Partner beteiligt sind. Für den letzten Fall wird in der Regel ein zentrales EDC-System von allen Beteiligten gemeinsam genutzt, wodurch gleichzeitig sichergestellt wird, dass durch die Verwendung identischer Fragebögen auch ein einheitliches Datenmodell genutzt wird.

In vielen Fällen überlappt die in einem EDC-System angestrebte Datenerhebung mit der Dokumentation, die in der klinischen Versorgung mittels der elektronischen Krankenakte ohnehin durchgeführt werden muss. Um hier eine Doppelerfassung zu vermeiden, wird angestrebt, das EDC-System mit der elektronischen Patientenakte zu verknüpfen und so jeden Datenpunkt nur einmal zu erfassen [2]. Für eine derartige Kopplung sind die Anforderungen von Seiten des Datenschutzes hoch. Aus diesem Grund erfolgt die Anbindung eines EDC-Systems in den meisten Fällen nur dann, wenn es innerhalb der gleichen Einrichtung wie das Krankenaktensystem betrieben wird.

Um auch in einem multizentrischen Szenario die datenschutzkonforme Anbindung von EDC-Systemen an die lokalen Krankenaktensysteme zu ermöglichen, schlagen wir eine Föderierte Architektur für EDC-Systeme vor (federated EDC, kurz: fEDC). Dabei wird in jeder beteiligten Einrichtung eine Instanz des fEDC-Systems unter Einhaltung der lokalen Datenschutzregeln (z.B. innerhalb eines Kliniknetzwerks) installiert. Dies ermöglicht einen lokalen Datentransfer vom EDC-System in die elektronische Patientenakte. Über das fEDC-Protokoll wird dann sichergestellt, dass die Formulardefinitionen zwischen den Partnern synchronisiert werden. Nach Abschluss der Studie werden die erhobenen Daten dann gemäß dem vereinbarten Studienprotokoll zusammengeführt und ausgewertet.

Um die grundsätzliche Machbarkeit des Ansatzes zu demonstrieren, wurde ein fEDC-Prototyp auf Basis der open-source Software openEDC entwickelt [3]. Insgesamt bietet die fEDC einen neuen Ansatz um die Integration von elektronischen EDC und Krankenaktensystemen auch in multizentrischen Szenarien zu ermöglichen. Dadurch wird die datenschutzkonforme Erhebung qualitativ hochwertiger Datensätze unterstützt.

Literatur

- [1] B. Walther, S. Hossin, J. Townend, N. Abernethy, D. Parker, D. Jeffries, Comparison of Electronic Data Capture (EDC) with the Standard Data Capture Method for Clinical Trial Data, *PLoS One* 6 (2011). <https://doi.org/10.1371/journal.pone.0025348>.
- [2] P. Bruland, C. Forster, B. Breil, S. Ständer, M. Dugas, F. Fritz, Does single-source create an added value? Evaluating the impact of introducing x4T into the clinical routine on workflow modifications, data quality and cost-benefit, *Int. J. Med. Inform.* 83 (2014) 915–928. <https://doi.org/10.1016/j.ijmedinf.2014.08.007>.

- [3] M. Ganzinger, M. Blumenstock, A. Fürstberger, L. Greulich, H.A. Kestler, M. Marschollek, C. Niklas, T. Schneider, C. Spreckelsen, E. Tute, J. Varghese, M. Dugas, Federated electronic data capture (fEDC): Architecture and prototype, *J. Biomed. Inform.* 138 (2023) 104280. <https://doi.org/10.1016/j.jbi.2023.104280>.

Section Classification in German Discharge Letters using Few-Shot Learning and Prompting

Phillip Richter-Pechanski^{a,b,c,d,e}, Philipp Wiesenbach^{a,b,d}, Dominic M. Schwab^b, Christina Kiriakou^b,
Mingyang He^{a,b,e}, Nicolas A. Geis^{b,d}, Anette FRANK^{e*}, Christoph Dieterich^{a,b,c,d*1}

^a*Klaus Tschira Institute for Computational Cardiology, Heidelberg, DE*

^b*Department of Internal Medicine III, University Hospital Heidelberg, DE*

^c*German Center for Cardiovascular Research (DZHK) - Partner site Heidelberg/Mannheim, DE*

^d*Informatics for Life, Heidelberg, DE*

^e*Department of Computational Linguistics, Heidelberg University, DE*

To facilitate clinical routine, medical information extraction (IE) research aims to make use of unstructured information as found in clinical documents such as discharge letters. In order to do so, most supervised machine learning approaches rely on large amounts of manually annotated training data. However, recent developments in NLP show promising results in text classification tasks using pre-trained language models (PLM) and prompts [1], severely reducing the need of supervision.

In our work we evaluate pattern-exploiting training (PET, [2]), a semi-supervised few-shot learning method using prompts. Prompting exploits the ability of PLMs to infer knowledge from context without the cost-intensive effort of model fine-tuning on large training datasets. In combination with lightweight supervision they achieve state-of-the-art results on various text classification tasks.

Hence, we explore this method in a German clinical IE text classification task, where a model classifies sentences from German discharge letters from the cardiovascular domain into eleven section categories, using section annotations of a publicly available German clinical corpus.

We evaluate PET using eight differently pre-trained PLMs and compare their results to baseline models used in current state-of-the-art approaches by running experiments on various few-shot training sizes. Our best performing PET model is based on a domain- and task-adapted BERT-large model [3], [4]. It achieves strong improvements over the baseline and publicly available PLMs for small shot sizes. We show that the performance gain diminishes with increase in sample sizes – an expected behavior that reinforces the incentives of our experiments. By studying interpretability methods, we were able to optimize the efficiency of our few-shot training sets, which improved the performance of our BERT-base model to a point where it achieves comparable classification results to the BERT-large model for selected section classes.

Overall, our results imply that PET can significantly improve the cost-efficiency of clinical text classification tasks on low-resource languages like German, by considerably reducing the need of annotated data *and* computational resources. We hope that our findings will pave the way to simplify the development of accurate section classification models to support, e.g. the automatic extraction of patient medication information or cardiovascular risk factors from discharge letters in clinical routine.

- [1] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language Models are Unsupervised Multitask Learners," OpenAI blog, vol. 1, no. 8, p. 9, 2019.

¹ *Corresponding authors: phillip.richter-pechanski@med.uni-heidelberg.de, christoph.dieterich@med.uni-heidelberg.de, frank@cl.uni-heidelberg.de. CD and AF share the last authorship.

- [2] T. Schick and H. Schütze, "Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference," *EACL 2021 - 16th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*, pp. 255–269, Jan. 2020, doi: 10.18653/v1/2021.eacl-main.20.
- [3] S. Gururangan *et al.*, "Don't Stop Pretraining: Adapt Language Models to Domains and Tasks," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Jul. 2020, pp. 8342–8360. doi: 10.18653/v1/2020.acl-main.740.
- [4] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of the 2019 Conference of the North*, pp. 4171–4186, 2019, doi: 10.18653/V1/N19-1423.



Impulsvorträge

ENABLE - innovative care concepts in the treatment of breast cancer patients

Thomas M. Deutsch¹, André Pfob^{1,2,3}, Kathrin Haßdenteufel¹, Laura Bodenbeck¹, Thao-Vy Le¹, Christiane Breit¹, Ekaterina Martynenko¹, Fabian Riedel¹, Manuel Feißt⁴, Carlo Fremd², Katharina Smetanay², Laura Michel², Lea Vollmer⁵, Tobias Engler⁵, Andreas D. Hartkopf^{5,6}, Sara Y. Brucker⁵, Andreas Schneeweiss^{2,7}, Markus Wallwiener¹

¹ Department of Obstetrics & Gynecology, Heidelberg University Hospital, Germany

² National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ), Heidelberg, Germany

³ MD Anderson Center for INSPIRED Cancer Care (Integrated Systems for Patient-Reported Data), The University of Texas MD Anderson Cancer Center, Houston, USA

⁴ Institute of Medical Biometry, University of Heidelberg, Germany

⁵ Department of Women's Health, University of Tübingen, Tübingen, Germany

⁶ Department of Women's Health, University of Ulm, Ulm, Germany

⁷ German Cancer Research Center (DKFZ), Heidelberg, Germany.

Objective:

The clinical importance of interventions based on Patient Reported Outcomes (PROs) in cancer treatment has been impressively demonstrated, not least by the work of Ethan Basch (Basch et al., JAMA 2016). The multicenter ENABLE study digitally records via smartphone app both health-related quality of life (HRQoL) and stratified side effects of breast cancer patients under therapy at Heidelberg, Tübingen, and Mannheim University Hospitals. The aim of the study is the improvement of therapy adherence, the recognition and timely treatment of critical side effects, and the objectification of QoL in the course of different therapy strategies. In addition, the study prospectively validates ML-based smart algorithms for the prediction of clinically relevant adverse events and treatment interruptions.

Methods:

Since March 2021, patients have been included prior to initiation of systemic therapy for breast cancer regardless of treatment setting. After 1:1 randomization into an intervention and a control arm, QoL assessments are performed at six fixed time points during therapy using validated questionnaires. In the intervention group, a short weekly survey of QoL is additionally performed using a visual analog scale (EQ-VAS). In case of significant deterioration, therapy-associated side effects are queried in a graduated manner, the treatment team is informed and interventions are initiated. Via the smartphone app, the patients also receive an overview of treatment and diagnostic appointments as well as an educational with sound information about their disease and therapy side effects.

Results:

To date, 456 of 600 patients have been included in the study. First interim evaluations showed a very high compliance regarding the answering of the questionnaires. In the intervention group, 5537 EQ-VAS questionnaires were answered (24 per patient), on the basis of which 689 interventions were triggered using digital symptom queries. Regarding user satisfaction, the app achieved an excellent result with a mean SUS score of 87.7 (SD 10.19).

Summary:

The app-based ePROM can be used to assess both the safety of oncological therapy and the quality of life of oncological patients and enables symptom-based intervention. Results on intervention effects and validation of the intelligent algorithms are still pending.

Klinisches Informationssystem OPICAD zur Versorgungsoptimierung schwerstkranker Kinder aus aller Welt

Eva Maria Hartmann, Thomas Killmann - Friedensdorf International

Peter Haas – FH Dortmund

Abstract

Motivation: Die gemeinnützige Kinderhilfsorganisation Friedensdorf International holt jährlich ungefähr 200 schwerstkranke und verletzte Kinder aus Kriegs- und Krisengebieten zur medizinischen Behandlung nach Deutschland. Zwischen ihren Krankenhausaufenthalten werden diese in der Oberhausener Heimeinrichtung des Vereins betreut. Die medizinische Versorgung erfolgt durch Krankenhäuser bundesweit sowie im hauseigenen Medizinzentrum, welches über einen eigenen Eingriffsraum und ein Rehabilitationszentrum verfügt [1]. Die Dokumentation, Prozesssteuerung und Kommunikation in dieser verteilten Organisationsstruktur ist damit hochkomplex. Nachdem die Grenzen der bisher freitextbasierten Dokumentationssoftware erreicht waren, bestand die Notwendigkeit einer Migration. Dies war die Grundlage für das Projekt ophEPA2OPICAD, im Rahmen dessen ein webbasiertes prozessorientiertes klinisches System angepasst und eingeführt wurde.

Methodik: Die Anforderungsermittlung erfolgte durch Analyse des Altsystems sowie Workshops hinsichtlich Prozessabläufen, Organisationsunterstützung und Dokumentationsanforderungen mit bereichsspezifischen Fokusgruppen. Sodann wurde die Parametrierungs-, Änderungs- und Erweiterungskonzeption für das an der FH Dortmund seit 2005 kontinuierlich entwickelte generische Grundsystem ophEPA [2] und eine erste Umsetzung vorgenommen. Diese wurde mittels Usability Tests kombiniert mit dem standardisierten Fragebogen Post-Study System Usability Questionnaire (PSSUQ) [3] sowie weiteren Fragen zu Nutzerängsten evaluiert. Die endgültige ophEPA-Adaption wurde anschließend unter dem Namen ontology and phenomenon based international childrens aid documentation (OPICAD) implementiert und in regelmäßigem Austausch mit den zukünftigen Nutzer*innen im Friedensdorf weiter iterativ optimiert. Seit Anfang des Jahres 2023 befindet sich OPICAD im Echtbetrieb.

Ergebnisse: Durch Parametrierung und Zusatzimplementierung von speziellen Maßnahmendokumentationen steht ein umfassendes webbasiertes und prozessunterstützendes Enterprise System für das Friedensdorf zur Verfügung, wobei die Dokumentation strukturiert, semantikbasiert und vollständig papierlos erfolgt. An der Evaluation des Konzeptes haben 7 (von 30) zukünftige Nutzer*innen teilgenommen, die eine sehr heterogene Zusammensetzung in Bezug auf ihre Aufgaben und die zu nutzenden OPICAD-Funktionen hatten. Der PSSUQ zeigt eine hohe Zufriedenheit von 2,81 im Gesamtscore, 2,75 im Bereich Nützlichkeit des Systems, 2,48 in der Interface Qualität und 3,06 im Bereich der Informationsqualität. Im Usability Test zeigten sich anfängliche Schwierigkeiten bei der Durchführung der vorgegebenen Aufgaben aufgrund der Umstellung von einer reinen Freitext- auf eine strukturierte und objektbasierte Dokumentation.

Ausblick: Nach 6 Monaten Realbetrieb soll eine erneute Untersuchung der Nutzerzufriedenheit und der Güte der Prozess- und Dokumentationsunterstützung mit Hilfe des PSSUQs und an Hand einer Contextual Inquiry erfolgen.

References

- [1] Aktions Friedensdorf e.V., *Jahresbericht-2019*. [Online]. Available: https://friedensdorf.de/wp-content/uploads/2020/09/Jahresbericht-2019_final.pdf (accessed: Jan. 12 2023).
- [2] P. Haas, "Design und Implementierung einer WEB- und CDA-basierten einrichtungsübergreifenden Elektronischen Krankenakte," in *50. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds) 12. Jahrestagung der Deutschen Arbeitsgemeinschaft für Epidemiologie (dae)*, 2005.
- [3] J. R. Lewis, "IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use," *Int. J. Hum. Comput. Interact.*, vol. 7, pp. 57–78, 1995.

Entwicklung einer neuen Methodik zur automatisierten Bestimmung der Akzeptanz von Medikations-Alerts mittels Ereignisanalyse

Janina A. Bittmann^{1,2}, Camilo Scherkl², Andreas D. Meid², Walter E. Haefeli^{1,2}, Hanna M. Seidling^{1,2}

¹ Kooperationsseinheit Klinische Pharmazie, Universität Heidelberg, Im Neuenheimer Feld 410, 69120 Heidelberg

² Abteilung Klinische Pharmakologie und Pharmakoepidemiologie, Universitätsklinikum Heidelberg, Im Neuenheimer Feld 410, 69120 Heidelberg

In klinischen Entscheidungsunterstützungssystemen (Clinical Decision Support Systems, CDSS) integriert in elektronische Verordnungsmodule (Computerized Physician Order Entry, CPOE) wird die Akzeptanz von Medikations-Alerts je nach Systemarchitektur des CPOE-CDSS mit zwei Methoden gemessen: Die In-Dialog-Analyse kann genutzt werden, wenn ausgegebene Alerts durch die Anwender*innen aktiv überschrieben oder bestätigt werden müssen, während eine Ereignisanalyse gewählt wird, wenn die gesamte Medikationsdokumentation (nur indirekt) auf Anzeichen zur Umsetzung von Alerts evaluiert werden muss [1]. Letztere ist insbesondere bei manueller Durchführung zeitaufwendig; daher wurde ein automatisierter Ansatz zur Umsetzung einer Ereignisanalyse entwickelt und mittels Routinedaten aus dem CPOE-CDSS des Universitätsklinikums Heidelberg (UKHD) pilotiert.

Medikations-, Administrations- und Alertdaten von 3,5 Monaten aus dem CPOE-CDSS des UKHD (*i.s.h.med® Smart Medication*, Cerner Corporation bzw. *AiDKlinik®*, Dosing GmbH) wurden in 24-h-Zeitintervalle eingeteilt um automatisch einen Zusammenhang zwischen der Alertanzeige und potentiellen, dazugehörigen Verordnungsänderungen herzustellen. Die angezeigten Alerts wurden in aufeinanderfolgenden Zeitintervallen verglichen; wurden sie im nächsten Zeitintervall bzw. kontinuierlich bis zur Verlegung oder Entlassung angezeigt, kategorisierten wir sie als „persistierende Alerts“. Entfielen sie während gültiger und fortlaufender Arzneimittelverordnungen im nächsten Zeitintervall, wurde angenommen, dass die Verordnungsänderungen auf Grundlage der angezeigten Alerts basierten und sie deshalb als „ausbleibende Alerts“ definiert.

Ausgehend von 1670 Patientenfällen wurden 11428 Alerts analysiert, die im Median für drei Tage angezeigt wurden (Interquartilsabstand 1-7 Tage), wobei Alerts für Arzneimittel-Allergien am kürzesten (1 Tag) und Alerts für potentiell inadäquate Medikation bei Älteren (PIM) am längsten (4 Tage) angezeigt wurden. Am häufigsten wurden Alerts am ersten Tag des Krankenhausaufenthaltes getriggert (32%, N=3645). Insgesamt 44% der Alerts (N=5015) persistierten und 56% (N=6413) blieben im Verlauf aus, Alerts für Arzneimittel-Wechselwirkungen am häufigsten (81%, N=1915) und PIM-Alerts am seltensten (40%, N=199).

Die entwickelte automatisierte Methodik der Ereignisanalyse ist auf die Evaluation passiver, nicht-interruptiver Alerts abgestimmt, für deren Akzeptanzmessung bisher keine automatisierte Methodik bekannt war. Mit dieser Methodik könnten große Datenmengen longitudinaler Patientenfälle verarbeitet und die Verhältnisse persistierender und ausbleibender Alerts abgeleitet, perspektivisch verglichen und überwacht werden – eine solche Funktion könnte gleichermaßen zur weiteren Validierung der Technik, aber auch zum besseren Verständnis des Effekts von passiven, nicht-interruptiven Alerts verwendet werden, die insbesondere im nicht-englischsprachigen Raum vorrangig eingesetzt und vergleichsweise schlecht untersucht sind [2,3].

(350 /350 Wörter)

- [1] Hussain MI, Reynolds TL, Zheng K. Medication safety alert fatigue may be reduced via interaction design and clinical role tailoring: a systematic review. *J Am Med Inform Assoc* 2019;26(10):1141-49.
- [2] Carli-Ghabarou D, Seidling HM, Bonnabry P, et al. A survey-based inventory of clinical decision support systems in computerized provider order entry in Swiss hospitals. *Swiss Med Wkly* 2013;143:w13894.
- [3] Ploegmakers KJ, Medlock S, Linn AJ, et al. Barriers and facilitators in using a Clinical Decision Support System for fall risk management for older people: a European survey. *Eur Geriatr Med* 2022;13(2):395-405.

Medical Informatics Meets Medical Ethics

Ever since the advent of computers, the tendency to delegate tasks to machines has been prevalent also in the clinic. Artificial intelligence already helps medical staff with a multitude of different tasks, including precision dosing, predicting long-term therapeutic outcomes, and interpreting medical images. Genuinely ethical tasks have so far been excluded from automatisation.

With the COVID-19 pandemic, however, the need for the taking of thousands of morally relevant decisions within short time frames has arisen. Expanding the use of artificial intelligence into the realm of clinical ethics suddenly seems a worthwhile enterprise.

In the past months, our interdisciplinary team of engineers and ethicists have developed the world's first functional ethical advisory system for clinical application. Results were published as a Target Article in the American Journal of Bioethics and have sparked off a lively international debate (<https://www.tandfonline.com/doi/full/10.1080/15265161.2022.2040647>). Preliminary performance results are promising: the algorithm's recommendations do not deviate significantly from those of human ethicists. I will begin this talk with an analysis of the different moral frameworks on which an ethical advisor system could be based and explain how we used machine learning to incorporate Beauchamp and Childress' prima-facie principles. I shall show how we acquired suitable training data, designed the input categories and user interface, and captured the parameters of individual medical cases.

That one *can* do something, however, does not imply that one also *should*: the basic technological means now exist to aid bioethical decision-making; but should we really entrust this sensitive domain to artificial intelligence? Most people will find the prospect of autonomously driving vehicles taking morally relevant decisions in situations of unavoidable harm easier to accept than having judgments in clinical settings made by artificial intelligence – even if we assumed that they would result in comparatively great harms or benefits. I will therefore conclude by reflecting on the profound impact on medical ethics that new technologies of this kind are likely to have.

Der Vortrag kann wahlweise auf Deutsch oder auf Englisch gehalten werden!

Secure Distributed Medical Analytics for German Healthcare institutions using the Personal Health Train (PHT-meDIC)

Marius de Arruda Botelho Herr^{1,2,4}, Michael Graf⁴, Peter Placzek⁴, David Hieber⁴, Felix Bötte², Stephanie Biergans¹, Mete Akgün², Nico Pfeifer², Oliver Kohlbacher^{1,2,3,4}

¹Medical Data Integration Center, University Hospital Tübingen; ²Institute of Bioinformatics and Medical Informatics (IBMI), University Tübingen; ³Applied Bioinformatics, Dept. of Computer Science, University of Tübingen; ⁴Translational Bioinformatics, University Hospital Tübingen

Abstract

Transferring data between different hospitals is often restricted, and federated analysis of clinical data is a viable alternative. Current federated analytics frameworks (e.g., DataSHIELD¹ or MedCO²) are often limited in the type of input data or analysis that can be performed. In the Personal Health Train (PHT) paradigm, the analysis algorithm (wrapped in a 'train') travels between multiple sites (e.g., hospitals - so-called 'train stations'), hosting the data in their protected infrastructure, and only transfers results rather than the data. Structured pseudonymized clinical data is stored in FHIR servers at Data Integration Center's (DIC's) based on the HL7/FHIR profiles of the German National Core Data Set³. Implementing trains as secured containers enables even complex data analysis workflows to travel between sites, i.e., genomics pipelines or deep-learning algorithms - analytic methods that are generally not easily amenable. We present PHT-meDIC⁴, a productively deployed, interoperable, open-source implementation of the Personal Health Train paradigm. The scope of applications for this platform ranges from machine learning algorithms to sophisticated omics and image analysis with arbitrary input data. Light-weight virtualization permits the automated deployment of complex data analysis pipelines (e.g., genomics, image analysis) across multiple hospitals in a secure and scalable manner. We combine different open-source third-party services with several custom-developed services. A separation into various services allows flexible adaption and extension in a scalable form. We achieved constant monitoring and persistent execution of trains and are providing governance template documents for deployment. Hospitals have pseudo-identifiers within the infrastructure and can only access their repository, and such inference attacks are less likely. Results are always encrypted at rest. Only participating sites and the submitting user can access them. Manipulation of trains will be detected at any stage. Furthermore, researchers can use additional privacy mechanisms (e.g., Paillier cryptosystem). The execution is within an encapsulated environment using study specific FHIR servers or data warehouses. We successfully deployed the implementation for distributed analyses of large-scale data. Our platform has been extended for interoperability in the Leuko-Expert project with other Medical Informatics Initiative partners' PHT⁵ architecture. Documentation and source code is accessible at <https://github.com/PHT-meDIC>.

References

1. Marcon Y, Bishop T, Avraam D, Escriba-Montagut X, Ryser-Welch P, Wheeler S, Burton PB, González JR (2021) *Orchestrating privacy-protected big data analyses of data from different resources with R and DataSHIELD*. PLOS Computational Biology 17(3):e1008880. March 30, 2021 <https://doi.org/10.1371/journal.pcbi.1008880>
2. Jean Louis Raisaro, Juan Ramon Troncoso-Pastoriza, Mickael Misbach, Joao Sa Sousa, Sylvain Pradervand, Edoardo Missiaglia, Olivier Michielin, Bryan Ford, and Jean-Pierre Hubaux (2019) *MedCo: Enabling Secure and Privacy-Preserving Exploration of Distributed Clinical and Genomic Data*. IEEE/ACM Trans. Comput. Biol. Bioinformatics 16, 4 (July 2019), 1328–1341. <https://doi.org/10.1109/TCBB.2018.2854776>
3. Bild, R., Bialke, M., Buckow, K., Ganslandt, T., Ihrig, K., Jahns, R., ... Prasser, F. (2020). Towards a comprehensive and interoperable representation of consent-based data usage permissions in the German medical informatics initiative. *BMC Medical Informatics and Decision Making*, 20(1), 103. doi:10.1186/s12911-020-01138-6
4. de A. B. Herr, M., Graf, M., Placzek, P., König, F., Bötte, F., Stickel, T., ... Kohlbacher, O. (2022). *Bringing the Algorithms to the Data - Secure Distributed Medical Analytics using the Personal Health Train (PHT-meDIC)*. doi:10.48550/ARXIV.2212.03481
5. Sascha Welten, Yongli Mou, Laurenz Neumann, Mehrshad Jaberansary, Yeliz Ucer Yedieli, Toralf Kirsten, Stefan Decker, Oya Beyan; Privacy-Preserving Distributed Analytics Platform for Health Care Data. *Methods of Information in Medicine* 2022; (DOI:10.1055/s-0041-1740564)

Using DICOM headers in FHIR Database for enrichment of the Medical Informatics Initiative's core data set

Authors: Christian Bruns, Jan Maluche, Ralf Lützkendorf, Johannes Bernarding

Data Integration Center Magdeburg, Institute of Biometry and Medical Informatics, Medical Faculty, Otto-von-Guericke University, Magdeburg

Introduction

The MII core dataset defines basic (patients, encounter, diagnoses, procedures and laboratory values) and extension modules on how data are stored and made available(1). This dataset is enriched by additional information based on incremental extraction of DICOM tags from a XNAT API(2) with the goal to a better study recruitment and the building of cohorts. These extracted data were transformed to a FHIR resource and stored in a FHIR based data ware house(3).

Methods

We developed a spring batch-based approach, defining readers, processors and writers for the data(4). The reader calls a RESTful interface to assemble information about the XNAT internal structure (subjects, projects, experiments, scans). Based on XNAT internal modification timestamps, the job determines if resources needs to be created, updated or are already up-to-date. Two timestamps can be provided by XNAT: Firstly, when an experiment was added, secondly, the time if the dataset was changed, which is relevant for the mapping and interpreted in the reading process. The processor applies a modular mapping process which maps from dicom-tags to FHIR imagingstudy resource. This resource is forwarded to the writer, which stores the resource in a FHIR server including a timestamp, according to internal XNAT data.

The whole approach is packed in a one-shot docker container, which is highly customizable through configuration files. An internal dataset of 2TB MRI-scans from 3 different scanners were used for evaluation and compared to the results of linux4health/dicom-fhir-converter(5), which maps locally stored DICOM images to FHIR resources. For speed evaluation, the hole dataset is processed and timings are taken.

Results

In the quality evaluation it couldbe proven that the only differences within the JSON files are the additional extension and within the metadata. The incremental load could speed up the load by up to 97% in comparison to a batch process.

Discussion and Conclusion

By now, only basic information about the presence of images are processed as a proof of principle. By researcher's demand, the amount of information can be easily extended through the modular implementation. The deployed process can be easily included into an existing and site individual architecture and may be used for future clinical trials.

Acknowledgement

This work was funded in part by the German Federal Ministry of Education and Research (BMBF), Funding reference number: 01ZZ1801H. The funding body did not play any role in the design of the study and collection, analysis, and interpretation of data as well as writing of the manuscript.

Bibliography

1. [Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med. 2018 May;57\(S 1\):e50–6.](#)
2. [Herrick R, Horton W, Olsen T, McKay M, Archie KA, Marcus DS. XNAT Central: Open sourcing imaging research data. NeuroImage. 2016 Jan 1;124, Part B:1093–6.](#)
3. [HL7 FHIR v4.0.1 \[Internet\]. Available from: <https://www.hl7.org/fhir/>](#)
4. [Spring Batch: Overview \[Internet\]. Available from: <https://spring.io/projects/spring-batch#overview>](#)
5. [dicom-fhir-converter \[Internet\]. Linux For Health; 2021. Available from: <https://github.com/LinuxForHealth/dicom-fhir-converter>](#)

ImiEDC Pilot Implementation at the Department of Obstetrics and Gynecology Heidelberg

André Pfob^{1,2,3}, Christian Niklas⁴, Max Blumenstock⁴, Thomas M. Deutsch¹, Fabian Riedel¹, Jörg Heil¹, Martin Dugas⁴, Markus Wallwiener¹

¹ Department of Obstetrics & Gynecology, Heidelberg University Hospital, Germany

² National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ), Heidelberg, Germany

³ MD Anderson Center for INSPIRED Cancer Care (Integrated Systems for Patient-Reported Data), The University of Texas MD Anderson Cancer Center, Houston, USA

⁴ Institute of Medical Informatics, Heidelberg University Hospital, Germany

(350 word max)

Objective:

ImiEDC is a software that enables digital patient anamnesis and structured documentation of clinical data. We conducted a pilot implementation of ImiEDC at the Department of Obstetrics and Gynecology Heidelberg to evaluate the feasibility of a structured, digital anamnesis.

Methods:

The ImiEDC system was used to create a digital anamnesis questionnaire tailored to patients presenting with ovarian lesions. As of December 21st 2022, patients presenting with an ovarian lesion were asked to complete the digital anamnesis questionnaire on an iPad including reason for visit, medical history, and current symptoms.

Results:

Within the first month of the pilot implementation, 26 patients with an ovarian lesion presented at our hospital. Fifteen patients (57.7%) successfully completed the questionnaire and the digital anamnesis was transferred into the electronic data system. The main reasons for not using the digital anamnesis questionnaire were (1) two or more patients presenting at the same time (only one iPad available) and (2) patient wish to use paper-based questionnaire.

Summary:

The ImiEDC system was successfully piloted to enable a structured, digital anamnesis for patients presenting with ovarian lesions. Future evaluations with respect to workflow implications and clinical outcomes are planned.

Webanwendung zur Unterstützung einer SNOMED CT-basierten Postkoordination

Tessa Ohlsen¹, Cora Drenkhahn², and Josef Ingenerf³

¹ Institut für Medizinische Informatik, Universität zu Lübeck, tessa.ohlsen@student.uni-luebeck.de

² IT Center for Clinical Research, Universität zu Lübeck, c.drenkhahn@uni-luebeck.de

³ Institut für Medizinische Informatik, Universität zu Lübeck, josef.ingenerf@uni-luebeck.de

Mit der Nationallizenz steigen Interesse und Verwendung von SNOMED CT in Deutschland. Die Postkoordination von Konzepten wird aufgrund ihrer Komplexität und mangelnder Softwareunterstützung vielfach gemieden, obwohl gerade diese SNOMED CT zu einer flexiblen und mächtigen Interlingua macht. Daher wurde eine Webanwendung entwickelt, die die Erstellung von postkoordinierten Ausdrücken (PCE-Ausdrücken) unterstützt und die Einhaltung der von SNOMED CT spezifizierten syntaktischen und semantischen Regeln (Compositional Grammar und Concept Model [1]) sicherstellt.

Die Erstellung eines PCE-Ausdrucks erfolgt entweder nur auf Grundlage des Concept Models (Variante1) oder auf Grundlage eines Templates für einen bestimmten Sachverhalt (Variante2). Bei Variante1 werden für das eingegebene Fokuskonzept die Constraints laut Concept Model automatisch ermittelt. Der/die Benutzer:in wählt die relevanten Attribute, die im Concept Model definiert sind, sowie die Attributwerte basierend auf den dazugehörigen Wertebereichen. Bei Variante2 werden basierend auf dem Template (SNOMED CT Expression Templates [2] oder neu generiertes Template) die verwendeten Attribute angezeigt. Auf Grundlage des in dem Template definierten Wertebereiches eines Attributes wird ein SNOMED CT-Konzept als Attributwerte durch den/die Benutzer:in gewählt. Bei beiden Varianten erfolgt eine automatische Überprüfung der Eingaben. Bei Korrektheit wird der PCE-Ausdruck basierend auf der Compositional Grammar erstellt. Eine Speicherung in einem FHIR CodeSystemSupplement ist möglich. Die Templategenerierung ist ähnlich zu Variante2. Neben den Attributnamen und Wertebereichen werden zusätzlich die Kardinalitäten für die einzelnen Attribute sowie RoleGroups festgelegt. Die Speicherung des Templates erfolgt im JSON-Format.

Durch die Verwendung des Machine Readable Concept Model (MRCM), des Terminologieservers Ontoserver [3] sowie den FHIR-Operationen [4] wird eine unkomplizierte Implementierung ermöglicht. Der Ontoserver beinhaltet eine komplette Implementierung der Expression Constraint Language und Postcoordinated Expressions.

Für die Validierung wurden 41 PCE-Ausdrücke aus den MIOs der KBV [5] verwendet. Diese wurden basierend auf dem Concept Model mit der FHIR-Operation „\$validate-code“ validiert. Bei sechs dieser PCE-Ausdrücke ließen sich formale Verstöße nachaufweisen, die der KBV mitgeteilt wurden. Alle PCE-Ausdrücke wurden auf Grundlage beider Varianten die PCE-Ausdrücke fehlerfrei nachgebildet. Das zeigt, dass es mit Hilfe der Webanwendung erfolgreich und zuverlässig möglich ist, PCE-Ausdrücke zu erstellen.

Quellen:

- [1] Ingenerf J; Drenkhahn C. Referenzterminologie SNOMED CT; Interlingua zur Gewährleistung semantischer Interoperabilität in der Medizin. Springer Fachmedien Wiesbaden, 2023.
- [2] SNOMED International. "Template Syntax DRAFT Specification". In: International Health Terminology Standard Development Organisation (2017).
- [3] Metke-Jimenez A et al. "Ontoserver: a syndicated terminology server". In: Journal of Biomedical Semantics 15 (2018), S. 9–24. doi: 10.1186/s13326-018-0191-z.
- [4] CSIRO. FHIR API. Ontoserver. url: <https://ontoserver.csiro.au/docs/5.0/api-fhir.html> (besucht am 19.12.2022).
- [5] KBV-BASIS-PROFILE. KBV. url: <https://mio.kbv.de/display/BASE1X0/KBV-Basis-Profile> (besucht am 18.12.2022).

Ontologiebasierte Erzeugung klinischer Dokumente

Dirk Hüske-Kraus, Maris Healthcare

Das Problem, im Rahmen der Patientenversorgung zu Kommunikations- und Dokumentationszwecken Dokumente (Befunde, OP-Berichte, Arztbriefe) zu erstellen, konnte die Medizininformatik bislang nicht suffizient lösen: Textbaustein- oder templatebasierten Verfahren mangelt es an Expressivität und Limitationen von Spracherkennung werden augenfällig: Schlechte Erkennungsraten bei Nichtmuttersprachlern, keine Plausibilitäts- oder Vollständigkeitskontrolle und häufig wiederholt zu diktierende Passagen mit geringer Varianz.

Andererseits erzeugen Natural Language Generation (NLG) Systeme erstaunlich gute Texte mit wenigen Eingabedaten, allen voran GPT mit ChatGPT [1]. Angesichts der lange im Raum stehenden Idee des NLG-Einsatzes für klinische Dokumente [2], erscheinen diese Systeme als Königsweg.

Allerdings leiden diese Systeme an strukturellen Problemen, die ihren Einsatz für den genannten Zweck als nicht gangbar erscheinen lassen:

- Ihre Korrektheit ist nicht ausreichend für gesundheitsrelevante und entscheidungsleitende Dokumente, auch aufgrund ihrer Neigung zu „hallucinations“ [3].
- Sie benötigen unrealistisch große Mengen annotierter Trainingsdaten, um die klinisch relevante Varianz der jeweiligen Anwendungsdomäne abbilden zu können.

Als Alternative wird ein Weg vorgeschlagen, der auf folgendem Ansatz (Suregen [4]) basiert:

- Der relevante Wirklichkeitsausschnitt wird in einer Ontologie spezifiziert, Existierende Klassen werden verfeinert und Abhängigkeiten zwischen Attributen definiert.
- Es wird eine graphische Benutzeroberfläche (GUI) entworfen, in der alle klinisch relevanten Sachverhalte dokumentiert werden können, Rückwirkungen von Sachverhalten auf Ontologieebene auf die GUI werden modelliert.
- Es werden Schemata spezifiziert, wie die dokumentierten Sachverhalte sprachlich auszudrücken sind.
- Eine Generierungskomponente erzeugt mittels dieser Schemata und der durch Dokumentation im GUI erzeugten Instanzen der Ontologie den Dokumententext.

Dieses Verfahren wird in einem kommerziell verfügbaren System eingesetzt und weist folgende Vorteile auf:

- Möglichkeit, Plausibilitäts- und Vollständigkeitskontrollen zu implementieren
- Syntaktisch korrekte und semantisch plausible Ergebnisse, auch bei Nichtmuttersprachlern
- Strukturierte Daten (für Qualitätssicherung, Benchmarking, Abrechnung, Forschung) fallen als Nebenprodukt an

Wesentlicher Nachteil ist hier, dass die Konfiguration nicht unaufwendig anwendungsspezifisch erfolgen muss.

Referenzen:

1. OpenAI: ChatGPT, <https://openai.com/blog/chatgpt/>

2. Hüske-Kraus, D. (2003). Text generation in clinical medicine—a review. *Methods of information in medicine*, 42(01), 51-60.
3. Ishii, Y., Madotto, Andrea. A., & Fung, Pascale (2022). Survey of Hallucination in Natural Language Generation. *ACM Comput. Surv*, 1(1).
4. Hüske-Kraus, D. (2003). Suregen-2: A shell system for the generation of clinical documents. In *Demonstrations*.



Poster

Mobile Apps for COVID 19: First Results of a Systematic Review of Reviews

Authors:

Marco Aigner ^{1, 2, 3}

Johannes Schobel ¹

Walter Swoboda ¹

Felix Holl ^{1,4}

¹DigiHealth Institute, University of Applied Sciences Neu-Ulm, Neu-Ulm, Germany

²Ruprecht Karl University of Heidelberg, Heidelberg, Germany

³Heilbronn University of Applied Sciences, Heilbronn, Germany

⁴Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig Maximilian University of Munich, Munich, Germany

Corresponding author:

Felix Holl, Neu-Ulm University of Applied Sciences, Wileystr. 1, 89231 Neu-Ulm, Germany;

felix.holl@hnu.de; +49 731 9762 1613

Keywords:

SARS-CoV-2; Mobile Applications; mHealth; Telemedicine; Systematic Review.

Background: Over the past two years, severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) has caused a global pandemic. In other endemic situations, mobile applications (apps) have been used successfully. [1]

Objective: This study aims to identify and review published reviews of mobile apps used in association with the coronavirus disease 19 (COVID 19).

Methods: We conducted a systematic review of reviews following the PRISMA statement from 2020.[2] We searched Medline and Embase for peer-reviewed reviews of COVID-19 mobile apps published between 01/01/2020 and 04/25/2022. Review quality was assessed by modified versions of the overview quality assessment questionnaire (OQAQ) and a measurement tool to assess systematic reviews, AMSTAR version 2.[3], [4] Reviews were categorically synthesized.

Results: Out of the initial 17,611 studies, 24 studies were eligible for the analysis. Publication dates ranged from May 2020 to January 2022. 54% (n=13) of the studies were published in 2021, 33% (n=8) in 2020, and three were published in 2022. Most reviews analyzed apps from the USA, the UK, and India. Apps from most of the African, Middle- and South American countries were not analyzed in the reviews. Categorization, based on similarity, resulted in four clusters (app overview, privacy and security, MARS rating and miscellaneous). Nine reviews provided an overview of currently available apps for COVID 19, six studies reviewed apps

regarding privacy and security, and five studies rated apps using the Mobile Application Rating Scale (MARS). Four reviews did not fit into any of the other categories, of which two analyzed studies on the effectiveness of COVID 19 mobile apps, one evaluated the usability of COVID 19 mobile apps, and one identified challenges in app adoption in neo-liberal societies.

Discussion: Our study provides a comprehensive high-level overview of 24 reviews of apps for COVID 19 during the first two years of the pandemic. Reviews on apps from African, Middle- and South American countries are required as current literature focuses on European and North American apps. Apps were mostly developed with governmental backgrounds which were reflected by consistently high quality according to MARS. Rigorous studies are needed to provide evidence of their effectiveness.

References

- [1] D. Tom-Aba, P. M. Nguku, C. C. Arinze, and G. Krause, "Assessing the Concepts and Designs of 58 Mobile Apps for the Management of the 2014-2015 West Africa Ebola Outbreak: Systematic Review," *JMIR Public Health Surveill* 2018;4(4):e68 <https://publichealth.jmir.org/2018/4/e68>, vol. 4, no. 4, p. e9015, Oct. 2018, doi: 10.2196/PUBLICHEALTH.9015.
- [2] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *International Journal of Surgery*, vol. 88, p. 105906, Apr. 2021, doi: 10.1016/J.IJSU.2021.105906.
- [3] A. D. Oxman and G. H. Guyatt, "Validation of an index of the quality of review articles," *J Clin Epidemiol*, vol. 44, no. 11, pp. 1271–1278, 1991, doi: 10.1016/0895-4356(91)90160-B.
- [4] B. J. Shea *et al.*, "AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both," *BMJ*, vol. 358, Sep. 2017, doi: 10.1136/BMJ.J4008.

Abstract Use Case ASIC



The SMITH consortium, as part of the German Medizin Informatik Initiative (MII), is home to various clinical Use Cases. One of these Use Cases is ASIC (Algorithmic Surveillance in Intensive Care), where acute respiratory distress syndrome (ARDS) was investigated based on routine clinical data. These data were processed and analysed by artificial intelligence. The results are sent to an App, which is intended to support physicians on the intensive care unit (ICU) diagnosing the potential incidence of this syndrome, still being described with a mortality rate of about 45% [1]. The Use Case ASIC engaged eight (8) University Hospitals that included 31 intensive care units across Germany. In a Quality Improvement Strategy (QIS), particularly the utility of a mobile application for the earlier diagnosis of ARDS and improved treatment guideline adherence was investigated. The QIS was conducted as a three-phase, stepped-wedge design over 21 months. It endured through the COVID-19 pandemic and concluded with data from ~15,000 patients, which greatly exceeded the initial data size expectations at the start of the project. The ASIC patient dataset includes over 120 parameters, among others ARDS diagnosis, tidal volume, end-inspiratory, positive end-expiratory, & driving pressures. The Data Integration Centres and IT departments associated with each of the University Hospitals worked together within the Use Case ASIC to assemble this national dataset. This collaboration pushed the standardization of German medical data forward towards the goal of complete interoperability between University Hospitals nationwide. Outside of the MIMIC-III database, the ASIC dataset is the largest non-US critical care dataset, which provides a unique opportunity to support the development of AI-based prediction methods.



Make Metadata relational again: Entwicklung einer Metadatenstruktur innerhalb des UMG-MeDIC

Autoren: Caroline Bönisch¹, Dorothea Keszyüs, Tibor Keszyüs

Einleitung: Im Rahmen der klinischen Datenerhebung werden zu den jeweiligen Daten auch die Metadaten, d.h. Metainformationen wie beispielsweise ihr klinischer Herkunftsort, erhoben und gesichert abgelegt. Respektive Anforderungen an diese Metadaten können unter anderem aus den FAIR Prinzipien [1] und Standards wie Dublin Core abgeleitet werden [2].

Innerhalb des Betriebs des Medizinischen Datenintegrationszentrums der Universitätsmedizin Göttingen (UMG-MeDIC) werden Metadaten ein grundlegender Wert für die Qualitätssicherung der Daten, zur Weiterverwendung in der Forschung beigemessen. Die identifizierten Metadaten umfassen dabei Informationen über Quellsysteme, Data Owner und Consent, sowie Informationen über die Datengüte.

Methodik: In der ersten Aufbauphase des UMG-MeDICs wurden Metadaten innerhalb der Extract-Transform-Load Prozesse in Form von JSON Objekten in eine CouchDB übertragen. Die Umstrukturierung des UMG-MeDIC und die kontinuierliche Weiterentwicklung der Prozesse resultierte in der Umstellung der Datenhaltung hin zu einer relationalen Datenstruktur. Um dieser Änderung Rechnung zu tragen, erarbeitet das vorliegende Konzept eine generische Datenstruktur zur Speicherung der Metadaten aller klinischen Primär- und Sekundärsysteme der UMG. Dafür wurde vorab die bisherige Datenstruktur des MeDICs betrachtet und mittels Expertengesprächen geprüft, wie die bereits extrahierten Metadaten in die neue Struktur einzuordnen sind. Weiterhin wurden zusätzliche Metadaten auf Basis früherer Anforderungsanalysen identifiziert und eingeordnet.

Ergebnis: Eine generische, relationale Struktur für die Ablage der Metadaten wurde entwickelt. Bereits vorhandene Metadaten zu klinischen Systemen, aus der vorherig genutzten Datenbank (in Form von JSON-Objekten) konnten übernommen werden. Weitere Metadaten, welche bisher nicht automatisch aus den Datenbeständen ausgelesen wurden, werden nun mittels angepasster ETL-Strecken ebenfalls extrahiert und befüllen die neuen Relationen. Die relationale Struktur zur Erfassung der Metadaten umfasst dabei jeweils eine Tabelle für den Metadaten-Typ (Consent, Qualität, Lizenz), eine für die spezifische Ausprägung des Metadatum, und eine für die Metadatenbeschreibung inklusive der Verknüpfung zu anderen Source-Relationen.

Diskussion: Die vorliegende Arbeit zeigt die Entwicklung einer relationalen Metadatenstruktur zur Speicherung gemäß spezifischer Anforderungen im UMG-MeDIC. Die Anforderungen an die Relationen ergeben sich dabei aus dem derzeitigen Betrieb und den Prozessen innerhalb des UMG-MeDICs und der UMG. Die skalierbare Struktur unterliegt kontinuierlichen Weiterentwicklungen und Anpassungen an neue Anforderungen, die eine Datenhaltung mit klinischen Massendaten beinhalten. Innerhalb der Struktur sind demgemäß Erweiterungen vorgesehen und mitbedacht.

¹ Corresponding Author: caroline.boenisch@med.uni-goettingen.de

Referenzen:

[1] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

[2] ISO. (2019) *ISO 15836-2:2019, Information and documentation — The Dublin Core metadata element set — Part 2: DCMI Properties and classes*. Retrieved from <https://www.iso.org/standard/71341.html>

Interactive Feasibility and Data Explorer for Clinical Data (FEDEX)

Sanketa Hegde^{1,2}, Harald Wilhelmi^{1,2}, Christoph Dieterich^{1,2,3}

¹*Department of Internal Medicine III, University Hospital Heidelberg, DE*

²*Section of Bioinformatics and Systems Cardiology, Klaus Tschira Institute for Integrative Computational Cardiology, Heidelberg, DE*

³*Informatics for Life, Heidelberg, DE*

ABSTRACT:

Dynamic visualization of structured patient data is essential for clinical study conduct and decision making. Physician scientists appreciate interactive graphical interfaces to rapidly explore large structured data volumes originating from clinical studies and/or routine data. We have identified particular needs for hypothesis generation, cohort building and data quality control. The new Feasibility and Data Explorer (FEDEX) is an extension of the DZHK Feasibility Explorer [1], by the integration of Medical Data Explorer (MEDEX) software [2].

The FEDEX software is implemented with a JavaScript front-end with Bootstrap v4, a Python application layer using the Flask framework, and a PostgreSQL database back-end. Our open-source project is available via GitHub URL: <https://github.com/dieterich-lab/medex/tree/Fedex>. The main layout gives the user several options for visualizing clinical data in various formats: tabular, aggregated statistics or a whole range of plots: Scatter plot, Barchart, Histogram, Box plot and Heatmap. All of the aforementioned features enable hypothesis building from data exploration. Outlier detection is made easy and can be integrated in a quality control step of any study. New features, which can be used for cohort building, include powerful filtering options that allows the user to specify data subsets that meet criteria such as certain age intervals or smoking. Given use and access consent, FEDEX also offers customized data and figure download on demand. Informative error handling ensures that the user fills in all required input fields in order to obtain the desired results.

We have tested the performance of FEDEX on the UK Biobank data set with approximately 500,000 patient health records over a wide range of numerical and categorical data features. Typical response times on standard hardware with an Intel Core i7-7600U CPU at 2.80GHz and 16 GB RAM, fall below 2-3 seconds. FEDEX imports data in CSV format and has been tested on DZHK cohort data and various HiGHmed data sets as well. We currently plan to implement it with MII Broad Consent data at our local Medical Data Intergration Centre (MEDIC). Taken together, FEDEX empowers physician to do data science on clinical data.

- [1] Kindermann, A., Stepanova, E., Hund, H., Geis, N., Malone, B., & Dieterich, C. (2019). MedEx - Data Analytics for Medical Domain Experts in Real-Time. *Studies in health technology and informatics*, 267, 142–149. <https://doi.org/10.3233/SHTI190818>
- [2] Scheel, H., Dathe, H., Franke, T., Scharfe, T., & Rottmann, T. (2019, September). A Privacy Preserving Approach to Feasibility Analyses on Distributed Data Sources in Biomedical Research. In *GMDS* (pp. 254-261).

Robustness of Electroencephalography-Biomarkers for Major Depressive Disorder – an Exemplary Study with Alpha Bandpower

Yasmin Hollenbenders^{1,2}, Christoph Maier³, & Alexandra Reichenbach^{1,2*}

¹Center for Machine Learning, Heilbronn University

²Medical Faculty Heidelberg, University of Heidelberg

³Medical Informatics, Heilbronn University

*Correspondence: alexandra.reichenbach@hs-heilbronn.de

Objective biomarkers for Major Depressive Disorder (MDD) are needed for diagnosis, monitoring, and treatment development. Electroencephalography (EEG) is studied for obtaining MDD biomarkers. However, results vary widely regarding the usefulness of specific features from EEG (de Aguiar Neto & Rosa, 2019). Especially the small sample sizes in many clinical studies lead to an over-estimation of diagnosis accuracy (Rakić, Cabezas, Kushibar, Oliver, & Lladó, 2020). The comparison of studies is hampered by the vast possibilities in methods. A robust biomarker for clinical application has not yet been identified.

Publicly available datasets contribute to resolving these conflicts. Some of the datasets may be combined to yield large and diverse datasets, or can be used for replication and comparative studies.

Alpha bandpower is a biomarker for MDD diagnosis with contradictory results (Cai et al., 2020; Hosseinifard, Moradi, & Rostami, 2013). The goal of the current study was to test its robustness across publicly available EEG datasets. We identified the differences between the datasets, matched the data, and varied systematically commonly used processing steps.

Two datasets with resting-state EEG from MDD patients (n=24/34) and healthy controls (n=29/30) with 13 corresponding channels were obtained (Cai et al., 2020; Mumtaz, 2016). Data was re-referenced and artifacts were removed. Subsequently, data was either not normalized, or z-transformed channel or subject-wise. Time-series were then split in 20s windows. Ten windows from each participant and channel were selected randomly and their median alpha bandpower was used as robust feature. For each variation of processing steps, we trained four classifiers (linear SVM with subject-wise 6-fold crossvalidation). Two classifiers operated on the two datasets separately to diagnose MDD. Two classifiers tried to separate the two datasets.

Classification accuracy depends on normalization ($F_{2,60}=15.604$; $p<.001$) and classifier ($F_{3,60}=3.825$; $p=.014$). For the diagnoses in the different datasets the non-normalized data performs better across datasets ($F_{2,30}=4.184$; $p=.025$), which stems from only one dataset ($F_{2,30}=9.036$; $p<.001$). For the differentiation between datasets, normalization has a marked effect ($F_{2,30}=12.551$; $p<.001$).

Our results demonstrate that the normalization procedure, a common step in machine learning, has contradictory impact on the datasets. However, for combining datasets, normalization is a necessary procedure.

Cai, H., Gao, Y., Sun, S., Li, N., Tian, F., Xiao, H., . . . Zhao, Q. (2020). Modma dataset: a multi-modal open dataset for mental-disorder analysis. *arXiv preprint arXiv:2002.09283*.

de Aguiar Neto, F. S., & Rosa, J. L. G. (2019). Depression biomarkers using non-invasive EEG: a review. *Neuroscience & Biobehavioral Reviews*, 105, 83-93.

- Hosseinifard, B., Moradi, M. H., & Rostami, R. (2013). Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Computer Methods and Programs in Biomedicine*, 109(3), 339-345.
- Mumtaz, W. (2016). MDD Patients and Healthy Controls EEG Data (New). figshare. *Dataset. MDD Patients and Healthy Controls EEG Data generated by <https://doi.org/10.6084/m9.figshare.4244171>*, v2.
- Rakić, M., Cabezas, M., Kushibar, K., Oliver, A., & Lladó, X. (2020). Improving the detection of autism spectrum disorder by combining structural and functional MRI information. *NeuroImage: Clinical*, 25, 102181.

Developing a Mobile Serious Game Platform to Improve Orofacial Myofunctional Therapy for Children

Maximilian Karthan^{1,2}, Daniel Hieber¹, Rüdiger Pryss², Johannes Schobel¹

¹Institute DigiHealth, Neu-Ulm University of Applied Sciences

² Institute of Clinical Epidemiology and Biometry, University of Würzburg

Email: maximilian.karthan@hnu.de

Orofacial myofunctional disorders (OMDs) refer to a collection of conditions affecting oral and facial muscles, causing either functional or cosmetic issues. These disorders encompass improper posture of orofacial muscles, unusual chewing, swallowing, or speech patterns, dental misalignments, and blocked nasal passages (Russel 2017). The exact incidence of OMDs in children is difficult to determine, as they often go unrecognized, and, therefore, are often untreated. There is limited research on the prevalence of OMD in the general population, and the results vary greatly between studies. However, according to (Kellum 1994), the accepted prevalence of OMD is estimated to be 38%.

While a large part of the orofacial myofunctional therapy takes place together with speech therapists in offices, a not to be underestimated part of the therapy consists of exercises for home. A whistle can serve as a tool, resulting in activation of muscles near the lips and cheeks (Shah 2021). Unfortunately, these exercises are often not properly carried out by children or monitored by parents.

This poses two major problems: First, it is difficult for the therapists to assess how conscientiously the exercises have been carried out. Second, the exercises are not fun and parents often have to convince their children to practice.

Assessing, how often exercises were properly carried out, would enable therapists to tailor face-to-face sessions to children's current progress. Therapists assume that the therapy is more successful if the exercises are performed more often. There are also a lot of studies that show, that children are more likely to exercise if they have fun doing so (Watson 2016).

Therefore, we propose a mobile serious game platform to improve orofacial myofunctional therapy for children. It offers various games (e.g., Flappy Bird, a top-down car-racing game), which are not controlled via regular touch input, but the audio signal of before mentioned whistles. Therapists have the possibility to set weekly goals and track the progress of children. Finally, children may benefit from this approach, as they carry out their exercises while playing. Currently, we are designing a pre-study to evaluate the mobile application with children suffering from OMD.

References:

Russell, T., & Council, R. O. (2017). Orofacial Myofunctional Disorders.

Kellum, G. D. (1994). Overview of Orofacial mycology. Orofacial myology: Beyond tongue thrust. Rockville. MD: American Speech-Language-Hearing Association.

Watson, L. A., Baker, M. C., & Chadwick, P. M. (2016). Kids just wanna have fun: Children's experiences of a weight management programme. *British Journal of Health Psychology*, 21(2), 407-420.

Shah, S. S., Nankar, M. Y., Bendgude, V. D., & Shetty, B. R. (2021). Orofacial myofunctional therapy in tongue thrust habit: A narrative review. *International Journal of Clinical Pediatric Dentistry*, 14(2), 298.

Entwicklung einer digitalen Infrastruktur für eine intersektorale medizinische Versorgung - der klinische Anwendungsfall PICOS

Denise MOLINNUS^a, Angélique KURTH^a, Tim SEILER-WAGENHALS^a, Volker LOWITSCH^b, Gernot MARX^a und Johannes BICKENBACH^a

^a*Operative Intensivmedizin und Intermediate Care, Universitätsklinikum RWTH, Aachen, Deutschland*

^b*Healthcare IT Solutions GmbH, Deutschland*

Abstract

In Akutversorgungseinrichtungen (z.B. Intensivstationen von Krankenhäusern), aber auch in nachgelagerten Strukturen wie Reha-Zentren, Pflegeeinrichtungen oder Arztpraxen fallen große Mengen an medizinischen Daten an. Es gibt jedoch nur wenige Innovationen, die diese Ketten der longitudinalen Patientenversorgung miteinander verknüpfen und so einen verfügbaren Längsschnitt von Patientendaten erzeugen. Zwar gibt es erste Ansätze in Hochversorgungseinrichtungen wie Universitätskliniken, aber vor allem im außeruniversitären Bereich fehlt es an solchen Entwicklungen. Im Rahmen des Progress Hub DISTANCE (Digital Smart Hub for Advanced Connected Care) wird der interoperable Datenaustausch auf medizinische Einrichtungen der regionalen Versorgung ausgeweitet. Die übergreifende Aufgabe im Projekt DISTANCE ist es, die Pionierarbeit der Medizininformatik-Initiative (MII) zur Digitalisierung in der Medizin aus den Universitätskliniken in das außeruniversitäre Gesundheitssystem zu übertragen. Im Rahmen des Use Case PICOS wird eine patientenorientierte App entwickelt. Erstmals wird es möglich sein, eine Vielzahl von Vitalparametern als Längsschnittdaten bei den Patienten zu Hause zu erheben und den psychischen und physischen Zustand nach protraumiertem intensivmedizinischen Aufenthalt zu dokumentieren, die hochprävalent für ein sogenanntes PICS (post intensive care syndrome) anfällig sind. Langfristig werden die über die App erhobenen Daten anonymisiert, zusammengeführt und für Sekundärdatenanalysen und Forschungszwecke zur Verfügung gestellt. So entsteht ein strukturierter Überblick über verschiedenste elektronische Patientendatensätze von ehemaligen Intensivpatienten. Kliniker, Forscher, Informatiker und andere Mediziner an den Universitätskliniken können dann die gesammelten Daten nutzen, um durch die Bereitstellung und Analyse von Langzeitbeobachtungsdaten mit Methoden der künstlichen Intelligenz klinische Zustände und Behandlungsprozesse für bestimmte Patientengruppen zu optimieren.

Referenzen

- [1] Ramnarain D, Aupers E, den Ouden B, Oldenbeuving A, de Vries J, Pouwels S. Post Intensive Care Syndrome (PICS): an overview of the definition, etiology, risk factors, and possible counseling and treatment strategies. *Expert Rev Neurother.* 2021 Oct;21(10):1159-1177. doi: 10.1080/14737175.2021.1981289. Epub 2021 Sep 22. PMID: 34519235.

Entwicklung von ETL-Verfahren für die interoperable Nutzung von Gesundheitsdaten – Grundprinzipien und Lessons learned

E Thomas^{a,1}, HM Kruse^a, C Schubert^a, M Hoffmann^a, K Saleh^a, D Ammon^a

^a Datenintegrationszentrum, Universitätsklinikum Jena

Einleitung

Die Verwendung klinischer Routinedaten in der Forschung ist das wesentliche Ziel der Medizininformatik-Initiative[1]. Hierzu wurden Konzepte und technische Lösungen entwickelt, mit denen Daten in standardisierte Formate übertragen werden können. Daraus entstanden Data-Engineering-Aufgaben[2], da der ordnungsgemäße Einsatz von Standards und der vollständige Umfang der medizinischen Daten in bereits bestehenden klinischen Systemen nicht garantiert werden konnte. Hieraus wurden einige Erkenntnisse gewonnen, welche bei zukünftigen Entwicklungen als grundlegende Vorgehensweisen berücksichtigt werden können.

Methodik

Transformationen klinischer Routinedaten in standardisierte Datenformate sind eine Herausforderung, welche sich nicht nur durch ein Mapping verschiedener Attribute oder Datentypen lösen lässt. Die gewonnenen Daten sind sowohl für den Austausch zwischen Systemen als auch für übergreifende wissenschaftliche Auswertungen notwendig. Der Transformationsprozess muss sich bereits bei der Konzeption verschiedener ETL-Strecken an den häufig proprietären Datenformaten aus den medizinischen Primärsystemen orientieren. Der Einsatz offener Standards ermöglicht hingegen größere Freiheiten. Durch die Transformation kann zudem ein Umgang mit unvollständigen oder fehlerhaften Daten ermöglicht werden. Eingesetzte Werkzeuge für den Transformationsprozess müssen flexibel

¹ Korrespondenz-Autor, Eric Thomas, Universitätsklinikum Jena, Datenintegrationszentrum, Stoystraße 3, 07743 Jena; E-Mail: eric.thomas@med.uni-jena.de

genug sein, um die zur Verfügung stehenden Daten von einem gegebenen Datenformat in ein anderes zu übertragen und gegebenenfalls zu ergänzen. Dabei ist der korrekte Umgang mit Datentypen, Einheiten und standardabhängigen Konventionen wichtig, um Verfälschung der Daten zu vermeiden.

Ergebnisse

Für die Entwicklung verschiedener ETL-Strecken am DIZ Jena konnten folgende Grundprinzipien herausgearbeitet und an unterschiedlichen Beispielen dargestellt werden:

- Aufeinander aufbauende ETL-Verfahren
- Modulare Ansätze bei Verarbeitung unterschiedlicher Datenquellen
- Sicherstellung der Datenerhaltung aus den Quellsystemen
- Toleranz gegenüber fehlenden und fehlerhaften Informationen
- Nachträgliche Korrektur und Ergänzung der Informationen ermöglichen
- Rückführung der Erkenntnisse in den klinischen Betrieb
- Informationsanreicherung durch verschiedene Informationsquellen

Insbesondere die Auseinandersetzung mit den klinischen Quellsystemen ist notwendig, um Kompetenzen im Umgang mit den Datenformaten auszubilden. Hierdurch wurden auch Kenntnisse, die denen eines technischen Data Stewards entsprechen[3], erworben.

Diskussion

Ein wichtiges Ziel der Datentransformation sollte die Beibehaltung oder Verbesserung der Datenqualität aus dem klinischen Primärsystem sein. Durch die große Vielfalt an klinischen Primärdaten und unterschiedlichen Vorgehensweisen innerhalb der ETL-Strecken, muss für jede Transformation eine individuelle Transformationslösung erarbeitet werden. Für den Umgang mit Patientendaten ist hierbei jedoch die Beachtung grundlegender Vorgehensweisen für die Entwicklung notwendiger Transformationen, die sich aus den geschilderten Erkenntnissen ergeben, essenziell.

Referenzen

[1] Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. *Methods Inf Med.* 2018 Jul;57(S 01):e50-e56.

[2] Reis J, Housley M. Fundamentals of Data Engineering. Sebastopol, CA: O'Reilly Media; 2022

[3] Katzensteiner M, Müller A, Witte ML, Schewe N, Bott O (Leitungen der GMDS-Arbeitsgruppen "Datenmanagement in klinischen und epidemiologischen Studien" und "Curricula der Medizinischen Informatik"): Data-Steward(-ship) in der Digitalisierung des Gesundheitswesens. mdi: Forum der Medizin_Dokumentation und Medizin_Informatik 2022(4):123–125.

OpenEDC

Für interoperable und dezentrale Forschung

Motivation

Die Bedeutung von Daten wächst in unserer Gesellschaft stetig, so auch in der Medizin. Daten bilden die Grundlage für effektivere Behandlungen, die schnellere Markteinführung von Medikamenten und Medizinprodukten, bessere und schnellere Entscheidungen und effizientere Prozesse. Sie sind notwendig für jeden Fortschritt im Gesundheitswesen. Forschungsdaten in der Medizin können mit sogenannten Electronic Data Capture (EDC) Systemen erhoben und verwaltet werden [1]. In diesen Systemen laufen Daten aus unterschiedlichen Quellen zusammen, wie beispielsweise aus dem Labor, von Sensoren und aus Patientenbefragungen.

Problem

Das Hauptproblem bestehender EDC Systeme liegt in deren eingeschränkter Interoperabilität. Sie erzeugen Daten in herstellerspezifischen Formaten und behindern damit die Zusammenführung, gemeinsame Auswertung und sekundäre Nutzung von Informationen. Auch die Wiederverwendung von bereits existierenden Metadaten wird durch proprietäre Datenformate erschwert und führt zu einem weiteren Anstieg an heterogenen, inkompatiblen Forschungsdaten [2]. Darüber hinaus sind die meisten verfügbaren Systeme für Mediziner:innen nur schwer verfügbar, da sie entweder die Konfiguration auf einem Web-Server oder die Auslagerung von hochsensiblen klinischen Daten zu einem externen Dienstleister erfordern.

Lösung

OpenEDC ist ein quelloffenes EDC System welches auf dem neuen CDISC ODM 2.0 Standard (eine Veröffentlichung ist in Q2 2023 geplant [3]) sowie HL7 FHIR SDC [4] basiert. Darüber hinaus ist es frei öffentlich zugänglich und kann lokal auf dem eigenen Computer verwendet werden. Durch diese Kombination wird es allen Forschenden ermöglicht, standardkonforme Daten zu erheben und weiterzugeben. Neben klinischen Daten gilt das auch für Metadaten, um die erhobenen Daten interpretierbar zu machen. OpenEDC wird seit August 2022 im Rahmen des EXIST-Gründerstipendiums vom Bundesministerium für Wirtschaft und Klimaschutz gefördert.

[1] El Emam K, Jonker E, Sampson M, Krleza-Jerić K, Neisa A. The use of electronic data capture tools in clinical trials: web-survey of 259 Canadian trials. J Med Internet Res 2009 Mar 09;11(1)

[2] Habermann T. Metadata and Reuse: Antidotes to Information Entropy. Patterns 2020;1(1)

[3] ODM v2.0 Public Review Webinar, zuletzt aufgerufen am 16.02.2023 — <https://www.cdisc.org/events/webinar/odm-v2-0-public-review-webinar>

[4] FHIR Structured Data Capture, zuletzt aufgerufen am 16.02.2023 — <https://build.fhir.org/ig/HL7/sdc/>

Verteilte Analysen zur Untersuchung der aktuellen Dokumentation ausgewählter seltener Erkrankungen zwischen verschiedenen Krankenhäusern mit DataSHIELD

Raphael Verbüchel¹, Maria Weymann², Renata De Souza Falguera³, Peter Pallaoro³, Daniela Zöller², Stephanie Biergans¹

¹Medizinisches Datenintegrationszentrum (meDIC), Universitätsklinikum Tübingen, ²Institut für Medizinische Biometrie und Statistik (IMBI), Universitätsklinikum Freiburg, ³Klinikum rechts der Isar, Technische Universität München

Abstract

Die Medizininformatik-Initiative (MII), gefördert durch das Bundesministerium für Bildung und Forschung (BMBF), hat sich zum Ziel gesetzt Krankenhäuser, durch die Schaffung harmonisierter Gesundheitsdaten und einheitlicher Schnittstellen zu vernetzen. Das MII-Projekt "Collaboration on Rare Diseases" (CORD) soll dabei die Dokumentation seltener Erkrankungen unterstützen und in standardisierte Formate, wie das "Fast Healthcare Interoperability Resources" (FHIR) Format, überführen [1]. Aber auch Anwendungsmöglichkeiten von datenschutzkonformen Analysetools und Methoden im Bereich der seltenen Erkrankungen werden untersucht und eingesetzt. Beispielsweise kann das Konzept der verteilten Analyse hierbei genutzt werden, um Daten dezentral zu analysieren und die Weitergabe von ausschließlich anonymisierten (z.B. aggregierten) Analyseergebnissen sicherzustellen. Ein Softwaretool, das für verteilte Analysen entwickelt und sowohl in CORD als auch der MII getestet wird, ist DataSHIELD [2]. DataSHIELD integriert dabei notwendige Sicherheitsstandards, welche bei einer Analyse mit medizinischen Patientendaten benötigt werden. Ziel unserer Arbeiten ist die Durchführung einer verteilten Analyse für das Verbundprojekt CORD mithilfe von produktiv und für die Verarbeitung von Patientendaten freigegebenen DataSHIELD-Installationen und unter Beteiligung von mindestens drei Verbundpartnern aus verschiedenen Konsortien der MII. Bei der Umsetzung wurden in den drei Universitätsklinika Freiburg, München und Tübingen die technischen und organisatorischen Voraussetzungen geschaffen, um die Nutzung von DataSHIELD für die CORD spezifischen Fragestellungen sicherzustellen. Es wurden Skripte zum Durchführen der Analysen, zum Transformieren und Übertragen der Daten aus FHIR, oder anderen Formaten, in ein DataSHIELD-kompatibles Format erstellt. Die finale Analyse erfolgte mit einem zuvor definierten Datensatz und in enger Zusammenarbeit mit den klinischen Projektpartnern. Hierbei lag der Fokus auf der Kodierung von seltenen Erkrankungen durch ICD-10-GM-Codes und Orpha-Codes. Die Ergebnisse unserer Arbeiten zeigen die Herausforderungen von verteilten Analysen unter Nutzung von DataSHIELD im Kontext der MII und für den speziellen Anwendungsfall der seltenen Erkrankungen. Für die Analyse von Qualitätskennzahlen für die Kodierung von seltenen Erkrankungen an den Kliniken ist DataSHIELD nutzbar und liefert sinnvolle Ergebnisse. Dies eröffnet die Möglichkeit der Nutzung von DataSHIELD bei der wissenschaftlichen Begleitung der ab 2023 gesetzlich vorgeschriebenen Nutzung von ORPHA-Kodierung an den Kliniken.

References

1. Webseite von CORD in der MII, <https://www.medizininformatik-initiative.de/en/CORD>
2. Website des verteilten Analysetools DataSHIELD, <https://www.datashield.org/>

Function, Aim, and Merit of the Personal Health Train for the Secondary Use of Clinical Data in Research

Sascha WELTEN^a, Yongli MOU^a, Yeliz UCER YEDIEL^b, Oya BEYAN^{b,c},
Stefan DECKER^{a,b}, and Toralf KIRSTEN^d

^a *Chair Informatik 5, RWTH Aachen University, Germany*

^b *Fraunhofer Institute for Applied Information Technology (FIT), Germany*

^c *Institute for Biomedical Informatics, Faculty of Medicine and University Hospital, University of Cologne, Germany*

^d *Department of Medical Data Science, Leipzig University Medical Center, Germany*

The secondary usage of clinical data in research has been promoted and recommended by several experts and scientists during the last recent years [1]. However, current data protection regulations make seamless data usage difficult and research communities, especially in Germany, run the risk of missing a great opportunity to use this wealth of data [1].

Strongly related to these legal obstacles is the process of data access as it potentially poses threats to privacy and makes the usage of data more difficult [1,2]. Some access policies include either the prior download of the data or restricted access through a cloud, where data holders upload the data [2]. Nevertheless, both inherently contradict the data sovereignty of the data providers as they lose direct control over data, making traceability and provenance challenging. Further, the large data volumes of different formats silo-ed in each institution are not easily transferrable between the involved entities.

To circumvent these shortcomings and enable more responsible processing of data, the Personal Health Train (PHT) constitutes a valuable option to facilitate privacy-preserving data analysis [3]. The PHT follows the paradigm of Federated Learning, which reverts the workflow for the data analysis by bringing the analysis to the data. This methodology empowers data-holding institutions to stay in control over their data but makes data access possible for research. Beyond these advantages, the PHT facilitates the analysis of decentralised data, as it is not dependent on one specific programming language, data source technology, or fixed cryptographic protocols [3,4].

Recent research has shown that the PHT is agnostic to data standards (e.g. FHIR) and is capable of conducting data analysis on sensitive data distributed across multiple institutions [3,5]. In particular, these studies demonstrate that the PHT can manage complex analysis tasks and is a practicable solution for Machine Learning on distributed data.

As the call for the establishment of the secondary usage of data will gain traction during the upcoming years, solutions such as the PHT provide the infrastructure to enable clinical scientists to drive value from data and increase the outcomes of research.

References:

- [1] Jungkunz, Martin and Köngeter, Anja and Spitz, Markus and Mehlis, Katja and Cornelius, Kai and Schickhardt, Christoph and Winkler, Eva C., Forum Marsilius-Kolleg, Bd. 21 (2022): Stellungnahme zur Etablierung der sekundären Forschungsnutzung von Behandlungsdaten in Deutschland, 2022, <https://doi.org/10.11588/fmk.2022.1.91697>
- [2] Suver C, Thorogood A, Doerr M, Wilbanks J, Knoppers B, Bringing Code to Data: Do Not Forget Governance, J Med Internet Res 2020;22(7):e18087, URL: <https://www.jmir.org/2020/7/e18087>, DOI: 10.2196/18087
- [3] Oya Beyan, Ananya Choudhury, Johan van Soest, Oliver Kohlbacher, Lukas Zimmermann, Holger Stenzhorn, Md. Rezaul Karim, Michel Dumontier, Stefan Decker, Luiz Olavo Bonino da Silva Santos, Andre Dekker; Distributed Analytics on Sensitive Medical Data: The Personal Health Train. *Data Intelligence* 2020; 2 (1-2): 96–107. doi: https://doi.org/10.1162/dint_a_00032
- [4] Wirth, F.N., Kussel, T., Müller, A. *et al.* EasySMPC: a simple but powerful no-code tool for practical secure multiparty computation. *BMC Bioinformatics* 23, 531 (2022). <https://doi.org/10.1186/s12859-022-05044-8>
- [5] Welten, S.; Hempel, L.; Abedi, M.; Mou, Y.; Jaberansary, M.; Neumann, L.; Weber, S.; Tahar, K.; Ucer Yediel, Y.; Löbe, M.; Decker, S.; Beyan, O.; Kirsten, T. Multi-Institutional Breast Cancer Detection Using a Secure On-Boarding Service for Distributed Analytics. *Appl. Sci.* 2022, 12, 4336. <https://doi.org/10.3390/app12094336>