

Opinion Paper

Thomas Ganslandt* and Michael Neumaier

Digital networks for laboratory data: potentials, barriers and current initiatives

<https://doi.org/10.1515/cclm-2018-1131>

Received October 23, 2018; accepted November 7, 2018; previously published online November 29, 2018

Abstract: Medical care is increasingly delivered by multiple providers across healthcare sectors and specialties, leading to a fragmentation of the electronic patient record across organizations and vendor IT systems. The rapid uptake of wearables and connected diagnostic devices adds another source of densely collected data by the patients themselves. Integration of these data sources opens up several potentials: a longitudinal view of laboratory findings would close the gaps between individual provider visits and allow to more closely follow disease progression. Adding non-laboratory data (e.g. diagnoses, procedures) would add context and support clinical interpretation of findings. Case-based reasoning and disease-modelling approaches would allow to identify similar patient groups and classify endotypes. Realization of these potentials is, however, subject to several barriers, including legal and ethical prerequisites of data access, syntactic and semantic integration, comparability of items and user-centered presentation. The German Medical Informatics Initiative is presented as a current undertaking that strives to address these issues by establishing a national infrastructure for the secondary use of routine clinical data.

Keywords: electronic patient record; secondary use; systems integration.

Introduction

Medical care is increasingly being delivered by multiple providers across ambulatory and hospital settings, involving general practitioners and various diagnostic and therapeutic specialties. While integrated care initiatives are striving to coordinate this process [1], barriers often remain that impede communication between healthcare sectors and providers [2–4] and result in isolated shards of patient data collected by various providers along the care process. Patients themselves increasingly use wearables and connected diagnostic devices to acquire physiological measurements [5], leading to yet another separate trove of data that is not readily integrated with data acquired by professional healthcare providers. Paradoxically, even as more data is being acquired by more participants in the healthcare process, overall utility of these data is reduced by their fragmentation across multiple parties, sectors and technical platforms (Figure 1).

In this article, the potentials of digital networks to provide merged access to these fragmented data shards will be highlighted. Relevant barriers towards achieving integration and utilization of patient data will be shown, and efforts to address these issues will be presented based on the ongoing MIRACUM consortium project.

Potentials of digital networks for laboratory data

The “horizontal” integration of data from multiple health care providers offers a longitudinal view of laboratory findings which can “fill the gaps” between visits of individual providers and be supplemented with dense data points captured by the patients themselves, e.g. using wearables or connected diagnostic tests (Figure 2). This allows users to more closely follow disease progression and possibly avoid unnecessary repeat testing.

The “vertical” addition of data sources beyond the laboratory (e.g. diagnoses, prescriptions, procedures) can put laboratory data into context. It allows to identify

*Corresponding author: **Thomas Ganslandt**, Department of Biomedical Informatics of the Heinrich-Lanz-Center, Mannheim University Medicine, Ruprecht-Karls-University Heidelberg, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany, E-mail: thomas.ganslandt@medma.uni-heidelberg.de, <https://orcid.org/0000-0001-6864-8936>

Michael Neumaier: Institute for Clinical Chemistry, Mannheim University Medicine, Ruprecht-Karls-University Heidelberg, Mannheim, Germany

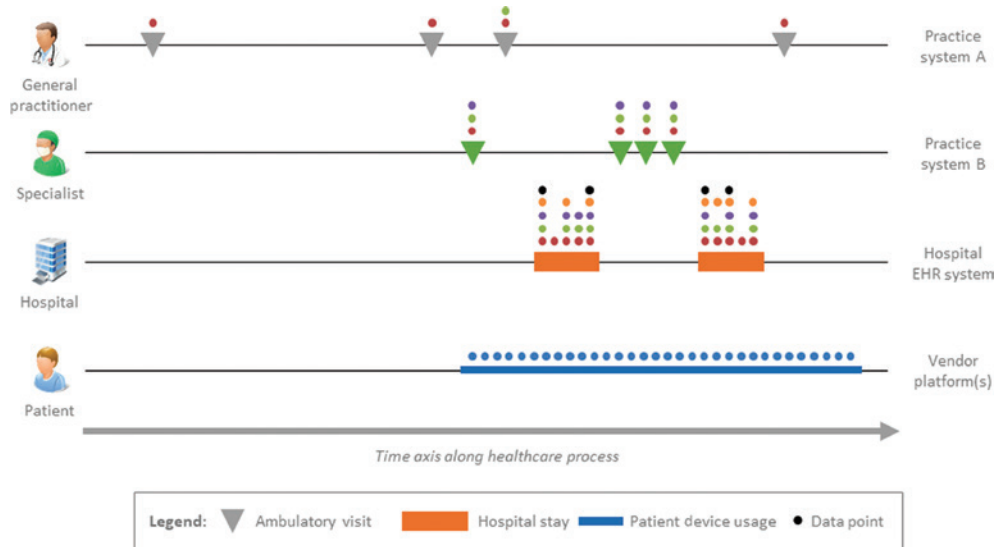


Figure 1: Fragmentation of data points along the healthcare process acquired by different providers in shared care as well as the patients themselves, and captured in different practice, hospital or vendor IT systems.

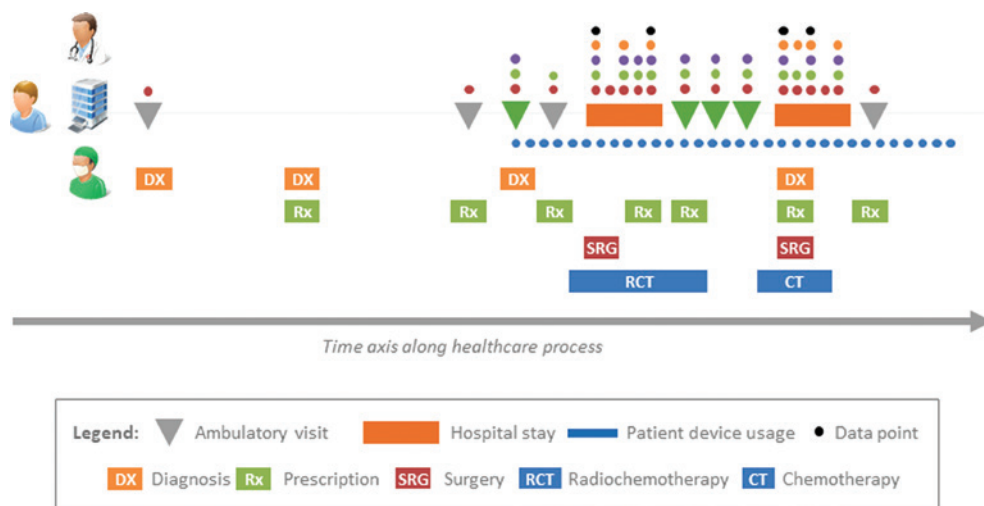


Figure 2: Longitudinal integration of patient laboratory data elements from collaborating providers on a "horizontal" axis as well as the "vertical" addition of context from additional diagnostic and therapeutic data sources.

relevant conditions, events or treatments that may influence laboratory measurements. Interpretation can thus take possible confounders into account. Apart from physician interpretation, the totality of integrated clinical data elements also enables the implementation of feasibility queries, clinical trial recruitment support and decision support platforms [6].

Two forms of decision support are exemplarily highlighted in Figure 3. In a case-based reasoning approach, data from an individual patient is used to find clusters of comparable patients, for which e.g. outcomes of different treatment options can be compared. In a multicenter scenario, this can include patient cohorts from other sites to

increase the available pool of eligible patients. Alternatively, data from a selected group of patients can be leveraged to generate disease models, e.g. by use of machine-learning approaches. The models can be applied to individual patient datasets to determine relevant endotypes, prognostic indicators or to support therapeutic decisions.

Barriers to implementation

In the following, several barriers towards achieving an integrated, cross-sectoral view of the electronic patient

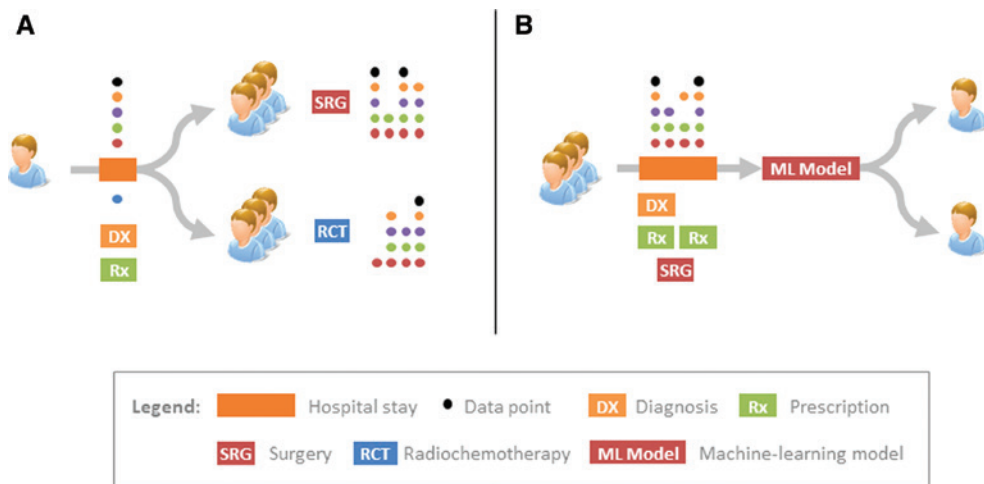


Figure 3: Secondary use of integrated healthcare data to (A) support case-based reasoning or (B) disease modelling approaches.

record are being described along the pathway from data access to data utilization.

Access to data

Even before questions of extracting and integrating data from fragmented sources can be addressed, a legal and ethical foundation granting access to these data sources needs to be established. Patients have the expectation that consent is obtained before data is used for healthcare or scientific purposes [7]. Studies have shown positive attitudes towards consent for data reuse both from clinical trial populations [8, 9] as well as general patient populations [10] when sufficient measures are implemented to protect patient privacy. However, healthcare delivery and administrative workflows should not be impeded by obtaining informed consent [11]. To this end, e-consent applications have been proposed that could serve to implement patient information and consent acquisition at least in part outside of administrative or clinical workflows [11].

When consent cannot be obtained, data re-use may still be possible based on research exemptions in locally applicable laws. In the United States, the Health Insurance Portability and Accountability Act (HIPAA) defines safe-harbor provisions for de-identified as well as limited datasets that can be used for research without requiring consent, including a defined set of identifiable attributes that need to be removed [12]. In Europe, unfortunately a more heterogeneous situation is the case. Even though an overarching General Data Protection Regulation (GDPR) has been introduced in Europe, variations persist as the process of implementation into national laws is still ongoing and permits local adaptations [13]. Rules may

also differ on a state-by-state level as is the case with data protection laws and provisions in hospital regulations in Germany [13]. This results in barriers towards implementation of multicenter projects across affected state or national boundaries.

Technical measures may also be applied to facilitate analyses on data. In distributed research networks (DRNs), analyses are carried out on data stored locally at healthcare provider sites and only aggregated results are made available [14]. This approach can be extended by applying cryptographical methods to carry out statistical analyses across distributed data subsets as implemented in the DataSHIELD platform [15]. Aggregated approaches preclude matching individual-level data across sites, which can impede analyses by erroneously including individuals multiple times or failing to correctly combine related data fragments into coherent datasets. This is especially relevant for analyses covering rare diseases or mobile patient populations accessing different healthcare providers. In this scenario, secure multiparty computing approaches can be applied to achieve privacy-preserving record linkage without divulging identifying information [16].

Syntactic and semantic integration of data

Once access to data sources has been established, data must be integrated and harmonized to permit cross-site analyses. Even within a single site, similar data may be documented in multiple systems, which need to be extracted and mapped to a common data structure [17]. Local coding schemes should be mapped to internationally established standard terminologies. For

the laboratory domain, the LOINC terminology (Logical Observation Identifiers Names and Codes) has been established both for use within the healthcare process as well as secondary research use [18–20]. SNOMED CT (Systematic Nomenclature of Medicine, Clinical Terms) can additionally provide value sets for specimens or non-numeric laboratory results (e.g. microbiological pathogens) and extensively covers many medical subject areas beyond laboratory concepts [21].

Secondary use of health data may occur on a “transactional” basis integrated into the healthcare delivery process. In this scenario, re-use focuses on accessing and/or transmitting individual patient records or analysis results within the context of patient care. Data formats like the HL7 CDA (Clinical Document Architecture) or HL7 FHIR (Fast Healthcare Interoperability Resources) can be used in this context to provide a common structured representation that can be transmitted through interfaces [22, 23].

In an “analytical” scenario, secondary use focuses on the large-scale analysis of data from patient cohorts. In this case, cross-site analysis can be facilitated by mapping local data elements to a common data model harmonized between participating sites. CDMs used on a broad scale have been developed in the OMOP OHDSI [24] and PCORNet [25] projects, among others.

Comparability of data items

Even when data items have been mapped to common data structures, terminologies and data models, it is not a given that records can be merged and analyzed together. In the laboratory domain, only a limited set of analytes is sufficiently standardized to ensure comparability [26]. Re-use of laboratory findings thus requires additional data items beyond the usual material, analyte, value, unit and reference range in order to support interpretation and comparability (e.g. method, test vendor, equipment and consumables batch numbers). These aspects cannot be addressed by LOINC coding of analytes, and are usually also not covered in standard data structures or CDMs used to represent laboratory findings in the research domain. LOINC coding of analytes may, however, support the implementation of projects to systematically analyze comparability of results between laboratories.

Usability and targeted presentation

While integration of data from multiple sources and longitudinally across the healthcare process increases the

breadth and availability of data, the adequate presentation of relevant subsets of data becomes a challenge of its own [27]. Both clinical users and patients need to be actively involved in the collection of requirements for their respective use cases [28, 29], and implementation should proceed iteratively to include structured user feedback on mockups, functional prototypes and deployed products. Tailored visual representations are an important tool to make complex constellations of data items accessible to users and highlight relevant attributes [30].

Current initiatives

The German Federal Ministry of Education and Research is currently funding the Medical Informatics Initiative (MII), a large-scale, long-term strategic project to establish a sustainable national infrastructure for the secondary use of routine clinical data, demonstrate its clinical utility and strengthen medical informatics as a discipline [31–33]. In the current phase from 2018 to 2021, four consortia are funded: DIFUTURE [34], HiGHmed [35], MIRACUM [36] and SMITH [37]. Each consortium will establish data integration centers (DIC) at its sites that will cover the extraction of data from local production IT systems, their integration into a coherent data warehouse, and the implementation of governance structures and processes to make these refined data available to local and external users. All consortia participate in a national steering committee and collaborate in working groups to ensure interoperability of data, platforms, processes and regulations across consortial boundaries. This includes the formulation of a national broad consent document in accordance with the state and federal data protection officers as well as the working group of German ethics committees. The goal is to prospectively obtain standardized, modular broad consent from patients at MII sites in order to enable data use beyond options provided by research exemptions in applicable laws. In order to ensure syntactic and semantic interoperability, all consortia collaborate in the development of an MII core dataset with detailed definitions of data structures and terminologies, based on established international standards [38]. Shared usage rules and governance processes are being developed to facilitate collaborative data use projects across institutional boundaries.

In the following paragraphs the approach of the MIRACUM consortium to address the abovementioned barriers and achieve the potentials of networked medical data will be presented in more detail. With 10 participating university hospitals and access to data from 12 million patients, MIRACUM is currently the largest MII consortium.

MIRACUM follows an agile approach with an early release of a minimum viable platform that is iteratively extended and optimized throughout the funding period, enabling it to adapt to user feedback and evolving requirements. A “MIRACOLIX” toolbox (Medical Informatics Reusable Ecosystem of Open source Linkable and Interoperable software tools) leverages the (re-)use of internationally established, freely available software platforms to foster sustainability as well as compatibility with related international research networks. The MIRACUM DIC architecture (Figure 4) implements the modular TMF data protection concept established in Germany for networked medical research [39]. A clinical data repository contains fully identified patient data and is made available locally within the treatment context to provide internal reporting and decision support on data extracted and integrated from routine source IT systems. A subset of data based on the MII core dataset is then harmonized to a common data model and stored in a second pseudonymized research data repository. The repository contains multiple data marts according data categories, including i2b2 (Informatics for Integration Biology and the Bedside) [40] and OMOP (Observational Medical Outcomes Partnership) [24] for clinical data, tranSMART for integration of molecular datasets [41] and XNAT for imaging data [42]. Research queries are subject to approval from a use-and-access committee as well as an ethics vote. Depending on project requirements and applicable consent or research

exemption, datasets are provided in an anonymized, project-specific pseudonymized or identified format. ID- and consent-management tools are applied to support this process.

MIRACUM is implementing three use cases to demonstrate the utility of the established infrastructure. Use case 1 (“Alerting in Care – IT Support for Patient Recruitment”) will support the recruitment of participants for clinical trials by leveraging routine data to provide candidate lists based on eligibility criteria stored in local trial registries. The second use case (“From Data to Knowledge – Clinico-molecular Predictive Knowledge Tool”) will apply machine-learning models on routine clinical and molecular data of asthma/COPD and neurooncology patients to identify endotypes and find comparable patient groups at other sites. The third use case (“From Knowledge to Action – Support for Molecular Tumor Boards”) aims to harmonize bioinformatics pipelines across MIRACUM sites, achieve integration of molecular and routine clinical data and tackle usability aspects of presenting the vast amount of complex data points required for annotation and presentation in an interdisciplinary tumor board.

To strengthen the discipline of medical informatics MIRACUM is establishing a joint master program “Bio-medical Informatics and Medical Data Science” as well as near-term focused training programs like summer schools and continued education programs that address staff currently attached to the consortium.

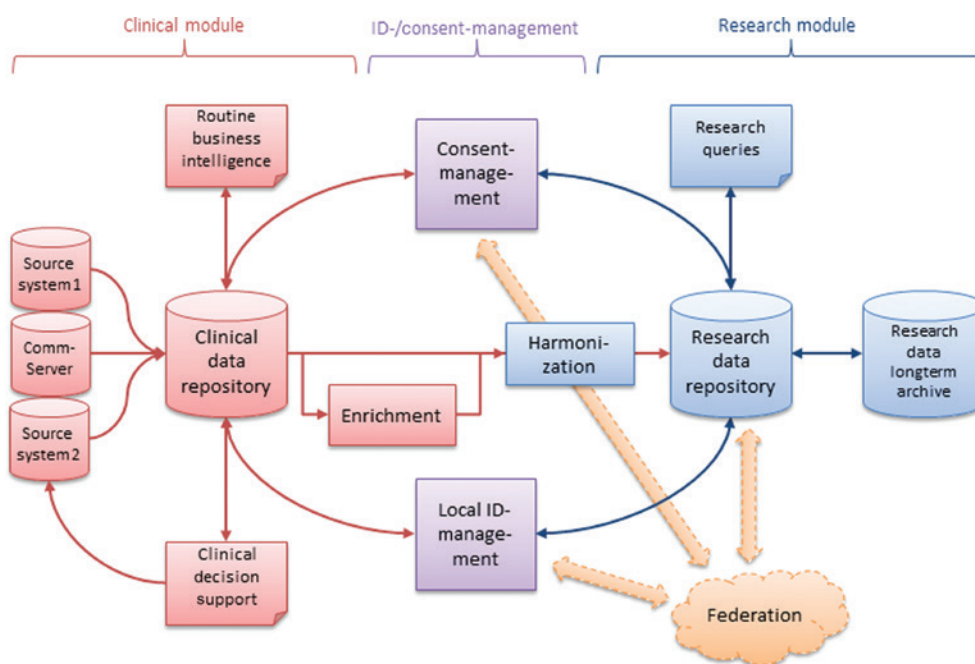


Figure 4: The MIRACUM data integration center (DIC) architecture.

Conclusions

Integration of electronic patient records currently fragmented across various providers and platforms into a coherent longitudinal dataset will enable significant potentials regarding holistic clinical interpretation, electronic decision support and novel visualization paradigms. Achieving access to data is complicated by varying regulations on an international as well national level. The establishment of a broad, modular patient consent harmonized across initiatives and possibly supported by electronic platforms to document and update patient consent can help to address this issue, but selection biases need to be taken into account (e.g. regarding patients unable to consent). Semantic harmonization of data elements to common syntactic structures and terminologies is an obligatory requirement for multicenter re-use of data. Internationally established standard terminologies like LOINC and SNOMED CT as well as standardized data structures like HL7 FHIR should be leveraged to achieve interoperability with collaborators. Apart from “post-hoc” mapping of existing data elements to such standards, it should be considered to establish “early mapping” of data items and value sets in routine clinical systems in order to harmonize already at the point of data capture. Clinical domain expert knowledge will be required in many cases to assess comparability of data items and participate in the interpretation of raw data and analysis results. Clinical and patient user requirements as well as usability feedback must be taken into account. The ongoing German Medical Informatics Initiative can serve as an example of an undertaking that tackles all major aspects of data access, interoperability, governance, clinical utilization and sustainability required to achieve the potentials of integrating large interdisciplinary medical datasets.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This work was funded in part by the German Federal Ministry of Education and Research within the Medical Informatics Initiative, Funder Id: 10.13039/501100002347, Grant ID 01ZZ1801E.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

- Schussel  Fillietaz S, Berchtold P, Kohler D, Peytremann-Bridevaux I. Integrated care in Switzerland: Results from the first nationwide survey. *Health Policy* 2018;122:568–76.
- Lang C, Gottschall M, Sauer M, K berlein-Neu J, Bergmann A, Voigt K. Interface problems between inpatient, GP and outpatient specialist care: viewpoint of general practitioners in Dresden. *Gesundheitswesen* 2018. doi:10.1055/a-0664-0470.
- Pain T, Kingston G, Askern J, Smith R, Phillips S, Bell L. How are allied health notes used for inpatient care and clinical decision-making? A qualitative exploration of the views of doctors, nurses and allied health professionals. *Health Inf Manag* 2017;46:23–31.
- Seeger I, Zeleke A, Freitag M, R hrig R. IT Infrastructure for biomedical research in north-west Germany. *Stud Health Technol Inform* 2017;243:65–9.
- Freckmann G, Mende J. Continuous glucose monitoring: data management and evaluation by patients and health care professionals – current situation and developments. *J Lab Med* 2018. doi:10.1515/labmed-2018-0119 [Epub ahead of print].
- Prokosch HU, Ganslandt T. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. *Methods Inf Med* 2009;48:38–44.
- Kim KK, Joseph JG, Ohno-Machado L. Comparison of consumers’ views on electronic data sharing for healthcare and research. *J Am Med Inform Assoc* 2015;22:821–30.
- Howe N, Giles E, Newbury-Birch D, McColl E. Systematic review of participants’ attitudes towards data sharing: a thematic synthesis. *J Health Serv Res Policy* 2018;23:123–33.
- Mello MM, Lieou V, Goodman SN. Clinical trial participants’ views of the risks and benefits of data sharing. *N Engl J Med* 2018;378:2202–11.
- Sanderson SC, Brothers KB, Mercaldo ND, Clayton EW, Antommaria AH, Aufox SA, et al. Public attitudes toward consent and data sharing in biobank research: a large multi-site experimental survey in the US. *Am J Hum Genet* 2017;100:414–27.
- Harle CA, Golembiewski EH, Rahmanian KP, Krieger JL, Hagmajer D, Mainous AG, et al. Patient preferences toward an interactive e-consent application for research using electronic health records. *J Am Med Inform Assoc* 2018;25:360–8.
- Standards for privacy of individually identifiable health information. Final rule. *Fed Regist* 2002;67:53181–273.
- Moln r-G bor F. Germany: a fair balance between scientific freedom and data subjects’ rights? *Hum Genet* 2018;137:619–26.
- Holmes JH, Elliott TE, Brown JS, Raebel MA, Davidson A, Nelson AF, et al. Clinical research data warehouse governance for distributed research networks in the USA: a systematic review of the literature. *J Am Med Inform Assoc* 2014;21:730–6.
- Wolfson M, Wallace SE, Masca N, Rowe G, Sheehan NA, Ferretti V, et al. DataSHIELD: resolving a conflict in contemporary bioscience – performing a pooled analysis of individual-level data without sharing the data. *Int J Epidemiol* 2010;39:1372–82.
- Laud P, Pankova A. Privacy-preserving record linkage in large databases using secure multiparty computation. *BMC Med Genomics* 2018;11(Suppl 4):84.
- Ganslandt T, Kunzmann U, Diesch K, P lffy P, Prokosch H-U. Semantic challenges in database Federation: lessons learned. *Stud Health Technol Inform* 2005;116:551–6.

18. Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leavelle D, Leland D, 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* 1996;42:81–90.
19. Zunner C, Bürkle T, Prokosch H-U, Ganslandt T. 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* 2013;20:293–7.
20. Bietenbeck A, Boeker M, Schulz S. NPU, LOINC, and SNOMED-CT: a comparison of terminologies for laboratory results reveals individual advantages and the lack of possibilities to encode interpretive comments. *J Lab Med* 2018 [Epub ahead of print].
21. Stearns MQ, Price C, Spackman KA, Wang AY. SNOMED clinical terms: overview of the development process and project status. *Proc AMIA Symp* 2001:662–6.
22. Dahlweid F-M, Kämpf M, Leichtle A. Interoperability of laboratory data in Switzerland – a spotlight on Bern. *J Lab Med* 2018 [Epub ahead of print].
23. Sabutsch S, Weigl G. Using HL7, CDA and LOINC for standardized laboratory results in the Austrian electronic health record. *J Lab Med* 2018 [Epub ahead of print].
24. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015;216:574–8.
25. Pletcher MJ, Forrest CB, Carton TW. PCORnet's Collaborative Research Groups. *Patient Relat Outcome Meas* 2018;9:91–5.
26. Panteghini M. Traceability, reference systems and result comparability. *Clin Biochem Rev* 2007;28:97–104.
27. Reichert D, Kaufman D, Bloxham B, Chase H, Elhadad N. Cognitive analysis of the summarization of longitudinal patient records. *AMIA Annu Symp Proc* 2010;2010:667–71.
28. Fylan F, Caveney L, Cartwright A, Fylan B. Making it work for me: beliefs about making a personal health record relevant and useable. *BMC Health Serv Res* 2018;18:445.
29. Hirsch JS, Tanenbaum JS, Lipsky Gorman S, Liu C, Schmitz E, Hashorva D, et al. HARVEST, a longitudinal patient record summarizer. *J Am Med Inform Assoc* 2015;22:263–74.
30. Zikmund-Fisher BJ, Scherer AM, Witteman HO, Solomon JB, Exe NL, Tarini BA, et al. Graphics help patients distinguish between urgent and non-urgent deviations in laboratory test results. *J Am Med Inform Assoc* 2017;24:520–8.
31. Haux R. Health information systems – from present to future? *Methods Inf Med* 2018;57:e43–5.
32. Gehring S, Eulenfeld R. German Medical Informatics Initiative: unlocking data for research and health care. *Methods Inf Med* 2018;57:e46–9.
33. Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. *Methods Inf Med* 2018;57:e50–6.
34. Prasser F, Kohlbacher O, Mansmann U, Bauer B, Kuhn KA. Data integration for future medicine (DIFUTURE). *Methods Inf Med* 2018;57:e57–65.
35. Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O, et al. HiGHmed – an open platform approach to enhance care and research across institutional boundaries. *Methods Inf Med* 2018;57:e66–81.
36. Prokosch H-U, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM: medical informatics in research and care in University medicine. *Methods Inf Med* 2018; 57:e82–91.
37. Winter A, Stäubert S, Ammon D, Aiche S, Beyan O, Bischoff V, et al. Smart medical information technology for healthcare (SMITH). *Methods Inf Med* 2018;57:e92–105.
38. Ganslandt T, Boeker M, Loebe M, Prasser F, Schepers J, Thun S, et al. Der Kerndatensatz der Medizininformatik-Initiative: Ein Schritt zur Sekundärnutzung von Versorgungsdaten auf nationaler Ebene [The medical informatics initiative core data set: a step towards the secondary use of routine clinical data on a national scale]. *mdi – Forum der Medizin, Dokumentation und Medizin-Informatik* 2017;20:17–21.
39. Pommerening K, Müller T. Leitfaden zum Datenschutz in medizinischen Forschungsprojekten: Generische Lösungen der TMF 2.0 [Recommendations for data protection in medical research projects: generic approaches of the TMF 2.0]. *Schriftreihe der TMF - Technologie- und Methodenplattform für die Vernetzte Medizinische Forschung e.V., Vol 11. Berlin: MWV Med. Wiss. Verl.-Ges; 2014.*
40. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010;17:124–30.
41. Athey BD, Braxenthaler M, Haas M, Guo Y. tranSMART: an open source and community-driven informatics and data sharing platform for clinical and translational research. *AMIA Jt Summits Transl Sci Proc* 2013;2013:6–8.
42. Marcus DS, Olsen TR, Ramaratnam M, Buckner RL. The extensible neuroimaging archive toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics* 2007;5:11–34.

Article Note: Invited opinion paper accompanying the talk given by Thomas Ganslandt at the 2nd EFLM Strategic Conference, 18–19 June 2018 in Mannheim, Germany (<https://elearning.eflm.eu/course/view.php?id=38>).