Review

Fried-Michael Dahlweid, Matthias Kämpf and Alexander Leichtle*

Interoperability of laboratory data in Switzerland – a spotlight on Bern

https://doi.org/10.1515/labmed-2018-0072
Received June 5, 2018; accepted July 10, 2018; previously published online September 4, 2018

Abstract: Laboratory data is a treasure chest for personalized medicine: it is – in general – electronically available, highly structured, quality controlled and indicative for many diseases. However, it is also a box with (probably more than) seven locks: laboratories use their own internal coding systems, results are reported in different languages (four official languages plus English with very distinct features in Switzerland), report formats are not uniform, standard nomenclature (e.g. Logical Observation Identifiers Names and Codes [LOINC]) is not routinely used and even these coding systems lack important information, including data, for example, about the specific "kit" used for testing or preanalytical procedures affecting the sample quality and result interpretability. Visualization of complex laboratory and reporting "-omics" data are additional challenges. Currently, there is no "passepartout" key for all these locks available, and also newer concepts such as Fast Health Interoperability Resources (FHIR) might not provide enough plasticity to unconditionally render laboratory data interoperable. In this short overview, we present current approaches in Switzerland with a specific focus on the exemplary Bernese implementations.

Keywords: clinical data warehouse; data exchange; laboratory data; Logical Observation Identifiers Names and Codes (LOINC); Switzerland.

Abbreviations and acronyms: AL, Analysenliste des Bundesamts für Gesundheit (BAG); CDA, Clinical Document Architecture; CDWH, clinical data warehouse; CSF, cerebrospinal fluid; EHR, electronic health record; ELGA, Elektronische GesundheitsAkte; EMR, electronic medical record; EPDG, Bundesgesetz über das elektronische Patientendossier; ETL, extract, transform, load; FHIR, Fast Health Interoperability Resources; Hadoop, the Apache Hadoop project; HL7, health level 7 standard; HL7 CDA, Clinical Document Architecture as part of the health level 7 standard; IDMS, isotope dilution mass spectrometry; IHE XD-LAB, Sharing Laboratory Reports (XD-LAB) Integration Profile of the "Integrating the Healthcare Enterprise (IHE) initiative"; L4CHLAB, LOINC for Swiss Laboratories; LIMS, laboratory information management system; LOINC, Logical Observation Identifiers Names and Codes; NLP, natural language processing; POCT, point-of-care testing; RDBMS, relational database management system; RELMA, REgenstrief LOINC Mapping Assistant; RIM, Reference Information Model; SAMS, Swiss Academy of Medical Sciences; SERI, State Secretariat for Education, Research and Innovation; SMART, Substitutable Medical Apps and Reusable Technologies; SNF, Schweizerischer Nationalfonds (Swiss National Science Foundation); SNOMED CT, Systematized Nomenclature of Human and Veterinary Medicine-Clinical Terms; SPHN, Swiss Personalized Health Network; SPREC, Standard PREanalytical Code; SQL, Structured Query Language; UI, user interface; UMLS, National Library of Medicine's Unified Medical Language System.

***Corresponding author: Alexander Leichtle,** University Institute of Clinical Chemistry, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland; IDSC – Insel Data Science Center, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland; and Directorate of Teaching and Research, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland, E-Mail: alexander.leichtle@insel.ch
Fried-Michael Dahlweid and Matthias Kämpf: Directorate of Technology and Innovation, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland; and IDSC – Insel Data Science Center, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

Introduction

In Switzerland, laboratory medicine is a rapidly developing discipline, not only in university hospital settings, but also for outpatient care. While in the last few years many efforts were focused on laboratory consolidation, process optimization and automatization, the now emerging big data science and data interoperability hold excellent promise for making efficient use of laboratory data generated in the course of clinical practice. Within

this digital revolution, laboratory medicine evolves more and more from an analytically focused "number delivery service" into a medical data generation and interpretation science along with all other scientific fields, which are now being entrained with the "digital revolution".

Looking into existing reality, the scientific community has seen a plethora of activities in biomedical and clinical data science collection, organization and analysis [1]. Also, in Switzerland, national initiatives have been launched, e.g. the action plan digitalization of the State Secretariat for Education, Research and Innovation (SERI[1]), the National Research Program "big data" of the Swiss National Science Foundation (SNF[2]), and the Swiss Personalized Health Network initiative of the Swiss Academy of Medical Sciences (SAMS[3]). All these programs and initiatives are intended to strengthen Switzerland's position at the forefront of "digital health" – and they are all utterly dependent on the interoperability of their data sources.

Within the healthcare ecosystem, laboratory medicine was an early adopter of digitalization, starting with electronic data capture and display in the 1980s, developing electronic process management and becoming increasingly paperless (not to mention a few inglorious exceptions). Laboratory data – and medical data in general – are complex, partly not perfectly structured and semantically complex. Exchanging such data requires specific normative or structural approaches, as failures in health-related information provision – particularly in laboratory and medication data – are a major contributing factor to medical errors. The US Joint Commission has introduced different patient safety goals, including test result communication guidelines [2]. Similar approaches have been introduced by many countries. In Switzerland, patient safety is one of the defined goals of the Swiss electronic health record law (Bundesgesetz über das elektronische Patientendossier [EPDG]) and will be in force by 2020.[4]

Today, laboratory reports are usually electronically available and highly structured. However, there are specific features that render the exchange of laboratory reports non-trivial:

First, laboratory results do not always consist of only numbers: there are tests with free text results, especially for complex diagnostics with results that would be uninformative for the treating physician, or there might be graphs (e.g. for cerebrospinal fluid [CSF] diagnostics, the so-called "Reiber" diagrams or flow cytometry scatterplots) or images (e.g. in hematology). A future challenge will be the transmission of "omics" results – including genomics, proteomics, metabolomics and future techniques. In these fields, a large amount of data are generated, and proper modes have to be found to transmit the relevant information in a clinically useful way [3], in Switzerland and beyond.

Second, additional information is needed for an adequate clinical interpretation of laboratory results. This could be preanalytical information about the patient or the sample [4–6], reference ranges for the specific analysis [7], additional information on the test kit or analytical instrument [8–10] or any other relevant "hidden" variable [11]. For the documentation of preanalytical variables, the Standard PREanalytical Code (SPREC) [12] was introduced and is currently in use at the BiobankBern.[5] It covers many relevant items; however, there is always a delay between the introduction of state-of-the-art technology for optimal sample pretreatment and its resemblance in the code reference (e.g. <2 h pre-centrifugation delay). In such cases, an unabridged documentation of all preanalytical steps – including the ones without a corresponding code – is necessary to be able to back match own data later on to an updated code version. The transmission of reference ranges together with the analyte concentrations is an issue of its own [13]: whereas a reference range on a laboratory report might be intended to support the physician in assessing a test result as "normal" or not, its suggested generation at each laboratory site renders it highly dependent on the hospital's patient cohort and thereby *per se* different between primary and tertiary care providers. Aside from questioning the concept of the monoparametric essence of the current reference ranges itself, reporting reference ranges always sets a patient in context with a "healthy" control group, which might or might not be relevant for the patient's condition or the scientific question of a collaborative project. Much more important seems the reporting of the test kit (vendor and version) used to generate the result, as different antibody specificities might yield different results for the same analyte, even if the same methodology is used. Actually, there is no standardized reporting for this kind of information in Switzerland. In contrast to these variables known to be important for the interpretation of laboratory test results, a large body of not yet correlated nor properly understood variables awaits being implemented and considered – the

1 https://www.sbfi.admin.ch/sbfi/de/home/das-sbfi/digitalisierung.html.

2 http://www.nfp75.ch/en.

3 http://www.sphn.ch.

4 https://www.admin.ch/opc/de/classified-compilation/20111795/index.html.

5 http://biobankbern.ch.

current reporting systems are neither thought nor dimensioned for such requirements: new concepts, e.g. graph-based semantics could offer solutions to this problem [14].

LOINC in Switzerland

In laboratory medicine, the Logical Observation Identifiers Names and Codes (LOINC) terminology became a universal coding system, despite its limitations. LOINC was introduced in 1994, and is maintained by a US non-profit organization called the Regenstrief Institute.[6] As of today, LOINC is being used in 172 countries across the globe and has expanded into additional medical fields, such as radiology, pharmacy and others. It is part of Clinical Document Architecture as part of the Health Level 7 standard (HL7 CDA), and thus a resource being able to be utilized in other coding systems. In its current version 2.63, released on December 15, 2017, it includes 86,528 terms, and many clinical laboratory systems support LOINC. Additionally, many, if not all, current electronic medical record (EMR) systems are also able to support LOINC. Within the USA, LOINC has been widely adopted by health information exchanges, reference laboratories, healthcare organizations, insurance companies, research applications and several national standards. The Department of Health and Human Services has adopted LOINC as the standard across federal agencies for laboratory result names, laboratory test order names and federally required patient assessment instruments. LOINC is a source vocabulary in the National Library of Medicine's Unified Medical Language System (UMLS) and was adopted by the National Cancer Institute's cancer Biomedical Informatics Grid through a formal review process [15]. Moreover, LOINC was adopted as the standard for laboratory orders and results as part of the Centers for Medicare and Medicaid Services Electronic Health Record (EHR) "Meaningful Use" incentive program in the Standards and Certification Criteria.

In Switzerland, LOINC is used to render laboratory data interoperable, and its country-wide implementation is by far more necessary as in Germany: Switzerland has four official languages, and four that are more or less regularly used in medicine throughout the country: German, French, Italian and especially in the scientific context English. The German "Kalium" is potassium in English and French, and potassio in Italian. Although widely in use, translation of standard terminology remains challenging, particularly regarding acronyms and abbreviations and linguistic syntax differences, which might cause specific difficulties in a multi-lingual country such as Switzerland. For multi-lingual purposes, the Regenstrief organization provides a software program called RELMA® (the REgenstrief LOINC Mapping Assistant) that helps in searching the LOINC database and mapping local terminology to LOINC terms. Although Switzerland has been amongst the first to adopt LOINC back in 2001, led by Jack Bierens de Haan at the Centre Suisse de Contrôle de Qualité, the current state is that LOINC translations in Swiss-national linguistic versions (Swiss French, Swiss German, Swiss Italian) lack additional support. Those vocabularies did not grow compared to other languages. Within the current RELMA® version, all three Swiss linguistic versions of LOINC remained at approximately 4900 translated terms, whereas the French (Canada) and the Italian (Italy) versions exceeded 40,000 terms, and the German (Germany) version 11,000 terms. As per the current Regenstrief policy, only translations of sufficient size (e.g. >10,000 terms) are enabled for searching in RELMA®, hence no Swiss linguistic version is currently searchable. One specific effort to cope with both the multi-linguistic challenges in Switzerland and the support of newer concepts, such as genomic information, is bundled in a Swiss Personalized Health Network (SPHN®) infrastructure development project called LOINC for Swiss Laboratories (L4CHLAB).[7] The availability of large amounts of data on these biological parameters is of major importance to research on precision medicine, and this particular project, which started in April 2018, aims to create the needed semantic interoperability.

LOINC has also been suggested to be mapped with the official Swiss reimbursement catalogue (Analysenliste AL[8]); however, the benefit – if restricted on this topic – would be at least limited: frequently, the cost of an analyte can be estimated by the technique used for measurement, and no further specification of the analyte would be needed for adequate accounting. At present, the codes of the Analysenliste are implemented throughout Switzerland, and mapping these with LOINC exclusively for reimbursement purposes would have to be justified by additional value and, if possible, without the restrictions inherent to the LOINC usage. LOINC is also suggested as a laboratory result coding system within the eHealth Suisse[9] environment (exchange format

6 https://loinc.org.

7 https://www.sphn.ch/en/projects/call-results-2017.html.

8 https://www.bag.admin.ch/bag/de/home/themen/versicherungen/krankenversicherung/krankenversicherung-leistungen-tarife/Analysenliste.html.

9 https://www.e-health-suisse.ch/fileadmin/user_upload/Dokumente/2018/D/180507_CDA-CH-LREP_de.pdf.

eLaboratoryReport – "Austauschformat eLaborbefund"), based on the Sharing Laboratory Reports (XD-LAB) Integration Profile of the "Integrating the Healthcare Enterprise (IHE) initiative" (IHE XD-LAB)[10] content profile. Although the wide adoption of LOINC has been a success, major challenges remain, particularly with the support of newer concepts, such as genomics [16], proteomics [17], metabolomics [18], and other omics-related ontologies. Since additional methodologies entered the world of healthcare laboratory data management, additional ontology-related hurdles also remain to be taken. For example, the sequencing of the human genome has led to a proliferation of innovative scientific research with respect to clinical medicine. This newly created clinical knowledge provides challenges for recording, managing and using the clinical data that result from that research. While many attempts have been demonstrated [16] to create terminology models to be utilized for recording of clinical genomic data, only LOINC offers meaningful content. As the LOINC concept is mainly single-analyte based, these techniques generating large datasets of correlated data items are not optimally suited to be represented within LOINC. Additionally, the preprocessing of the "omics" data contributes largely to its interpretation; therefore, all interactions with the data should likewise be stored, if not the "raw" files for later re-processing.

For microbiology testing, the usage of LOINC alone for data exchange seems additionally problematic – there are innumerous pathogens, sample retrieval sites and secondary derived materials (e.g. cultures) that impede a direct one-to-one mapping with specific LOINC codes. Amending the missing information using the Systematized Nomenclature of Human and Veterinary Medicine-Clinical Terms (SNOMED CT[11]) nomenclature seems to be a reasonable approach to ensure interoperability for this type of laboratory results. Similar is the situation for reporting somatic alterations in tumor tissues in pathological reporting. Ongoing work examines the applicability of mapping LOINC with SNOMED CT to better support laboratory information management systems (LIMS), EMR and other future implementations [19].

The laboratories are entitled to propose changes to LOINC by suggesting additional information that might be needed to specify a test, e.g. kit type or versions. Today, the Swiss university hospitals are using a highly homogenous infrastructure; however, joining data, e.g. with cantonal hospitals or outpatient laboratories, is difficult, e.g. tumor maker levels are not comparable despite a specific LOINC code (e.g. 83084-4 for CA19-9, regardless of vendor). For this problem, several approaches exist, such as introducing extension in FIHR (see below) or a specific mapping of LOINC with other reference systems, e.g. via the attribute "extended analysis information" in the Austrian Elektronische GesundheitsAkte (ELGA) value set.

With more and more specific and detailed number of LOINC codes, that is ever increasing, another problem arises: the back-merging. In our daily practice, we frequently see researchers requesting "creatinine". Depending on the purpose of their request, we should deliver different codes. If it is about measuring creatinine as a crude patient safety characteristic, we might report serum and whole blood creatinine from point-of-care testing (POCT), but if the request is about testing a new estimated glomerular filtration rate (eGFR) equation, we probably rely on centrally measured isotope dilution mass spectrometry (IDMS) calibrated plasma creatinine. Although LOINC now offers thematic LOINC groups,[12] this only partly solves the issues arising with thousands of unique LOINC codes in our data repositories: medical expertise of a laboratory specialist is anyway needed for valuable reporting, and providing clinicians and researchers with a non-curated self-service LOINC menu would be at least careless.

Fast Health Interoperability Resources (FHIR)

An advanced – although still emerging – concept to provide and exchange health information data is called Fast Health Interoperability Resources (FHIR).[13] Although many such attempts have been made in the field of medical and laboratory information management, and different standards have been implemented (e.g. HL7 v2, HL7 v3, Reference Information Model [RIM] and CDA), no such standard is able to comply with the ever-increasing complexity of digitization in healthcare. The main goal of FHIR is to ease health information interoperability without affecting the integrity of information. FHIR has seamless constructs for mapping HL7 RIM and other models, thus enabling strong ties to HL7's patterns and best practices without the need for in-depth knowledge of the RIM, or HL7 v3. However, current health information technology applications, such as EMR, or LIMS are productive systems and cannot be changed rapidly while they are still in daily use.

10 http://www.ihe.net/Laboratory/.
11 https://www.snomed.org/snomed-ct.
12 https://loinc.org/groups/.
13 https://www.hl7.org/fhir/.

Additionally, current data standards like HL7 v2 and v3 are less familiar, and far more complex as many modern and widely used web-based open standards. In order to comply with the need for such ultra-fast technology changeability, it became obvious to search for a widely supported mechanism to deploy such standards as quickly as possible.

One recent approach to such challenges became known as Substitutable Medical Apps and Reusable Technologies (SMART),[14] which incorporates FHIR. As outlined before, FHIR was designed to be standard based, open source and easy to implement, yet lacking the ability to integrate different health information applications. FHIR itself does provide definitions for ontologies, data models and data exchange methods, but not for accessing data, credentials or user interactions. To fill that gap, SMART has been defined to provide data profiles, standard authorizations, authentication technologies, as well as user interface (UI) integration patterns [20]. As of today, no known SMART on FHIR implementation has been published for LIMS, and the current level of LIMS interoperability, particularly in Switzerland, remains HL7 v2.x or CDA based. For future LIMS implementations, it might be paramount to evaluate the level of SMART on FHIR or FHIR readiness. Nevertheless, "plasticity" – as predicted by Christian Lovis [21] and described as transition of information technology to a semantic and temporal centric vision of information – will determine whether coding concepts will prevail or unlaureled descend in vain.

Laboratory data and clinical data warehouses (CDWH)

Recently, academic institutions across the globe have established significant efforts in the management of clinical and research data by enabling research data lifecycle as a means to support healthcare data acquisition, ingestion, curation, preservation, sharing and reuse [21, 22]. However, little is known as to what extent the design of so-called clinical data warehouses (CDWHs) or research data management platforms supports current and future requirements from a laboratory medicine point of view, particularly for multidimensional datasets. Additionally, many of those CDWHs focus on academic or research-related aspects, whereas their linkage with transactional clinical data management for time-critical day-to-day practice is missing.

To provide an overview, we describe existing research management implementations to draw conclusions for both related aspects of laboratory medicine, and particularly for Swiss implementation strategies. The latter will impact and likewise be impacted by national Swiss initiatives, such as the SPHN.

There are different approaches to organize a CDWH architecturally, and no gold standard has been found as of today. One very recent design approach seems to be promising and was considered in building the Inselspital's own CDWH: the requirements described by Nind et al. [23] are outlined in the following paragraphs. Additionally, this proposal is compared to the preliminary results derived from our own research.

1. Horizontal scaling: Most extract, transform, load (ETL) frameworks are adjusted for scaling vertically, which allows more records per time. The mechanism follows an approach to write-once per dataset. Data cleansing, data transformation and any optimizations will be done at hosted data only. However, this vertical scaling has limitations when it comes to rapid ETL and extraction of heterogeneous datasets, particularly mixing complex clinical measurements, complex laboratory data and genomic data. The solution seems to follow both vertical as well as horizontal scaling capabilities, especially in our case with several unconnected laboratory information systems (LISs), e.g. clinical chemistry and microbiology.
2. Data cleaning and curation tools: The current generations of research data platforms and CDWHs depend on the quality of pre-curated data and the reliability of ingested datasets. Additionally, those setups neither cope with retrospective rewriting nor with changing definitions and reference values, which is happening frequently, particularly regarding laboratory parameters. To avoid constant manual data clean-up and poor data quality, such cleansing and restructuring pipeline must be built into future research platforms and CDWHs. While currently the datasets from the laboratory exceed other source systems in data quality, new, patient-centered approaches, such as glucose self-monitoring or closed-loop devices, require a considerable amount of data curation (or at least the tagging of the source systems to enable both comprehensive and high-quality data queries).
3. Multifaceted extraction conversions: Current design approaches seem to lack the ability to facilitate multidimensional and frequent transformations. The latter is particularly important when it comes to optimizing and adapting datasets to feed into machine learning environments. At the Inselspital, we currently use

14 https://smarthealthit.org.

Markdown-based combined R and Structured Query Language (SQL) scripts to provide maximum flexibility together with thorough documentation.

4. Data lifecycle management: The current design follows a single data management resource approach, thus allowing access to many researchers. However, this conflicts with the approach to optimize cleansing and standardization of data. Hence, we implemented different data delivery options, from single file transfer of raw query outputs to highly customized reports including computations and graphics to meet the different requirements of researchers and clinicians.
5. Governance implementation: One of the high-level tasks in implementing a CDWH is the proper representation of institutional and study-related governance (cf. SPHN project "Development of a governance and quality management system for exchange of patient related data for research purposes."[7]). Without doubt, the different interests of data providers and/or data users are a challenge, which in our case is tackled by rigorous adherence to the permissions given in ethical waivers and an institutional board that assesses each request on the basis of intention of and fitness for purpose (as much as necessary, as little as possible). This pre-requires a working institutional governance scheme and – as governance information is not regularly implemented with the data – a large amount of curation.

Despite the broad availability of interoperability specifications, there is pervasive non-adherence to HL7 messaging standards (by using inappropriate levels), incomplete LOINC mapping (e.g. for not yet categorized analyses), inconsistently reported results (e.g. due to historical data structures) and non-adoption of SNOMED CT standards when reporting electronic laboratory results. These barriers must be overcome in order to efficiently deploy systems that utilize real-world data for secondary use. Ideally this is accomplished by source systems adopting and strictly adhering to specified standards. However, our experience shows that this is a cumbersome approach that depends on many, partly unforeseen instances including technical, legal and operational issues. Alternatively, complex systems can be used that compensate for sub-standard data by using robust techniques for mapping local codes to LOINC and multiple natural language processing (NLP) algorithms (cf. SPHN project "NLP-powered mapping of clinical reports onto SNOMED-CT concepts for tumor classification [NLPforTC]"[7]). Such systems also require periodic evaluation to ascertain accuracy. Today, it is evident that the accuracy of the source data is of utmost importance for the later use of this data in subsequent projects. However, it is at present not clear whom to charge with the efforts to cleanse databases that were improperly filled over the years.

The Insel clinical data warehouse (IDWH)

The first key principle of the Inselspital's CDWH initiative is to establish a data management and analysis platform not only for research, but also for patient treatment and hospital operations to optimally use synergies of a comprehensive analytics platform. The second key principle consists of building a modular architecture that is highly open and flexible, can evolve over time and thus will be able to meet future requirements. The core of the CDWH consists of a data lake infrastructure to maintain the data extracted from the source systems. Data persistence in the data lake is two-fold: a relational database management system for structured data and the Apache Hadoop project (Hadoop) distributed file system for unstructured and high-volume data. As we have an in-hospital system, we can link the data with the patient identification number, and technological bridges allow to query from both the structured and the unstructured part of the data lake. Our system scales horizontally to integrate various data sources that not only differ in technology and formats, but could also originate from either internal or external systems, admitting that part of our hospital infrastructure has an "external" label as it belongs to the university as a separate entity and not to the hospital – a rather common setting in university medicine. The preferred way to integrate a new source is to directly attach the source database to the ETL process and to integrate the raw source data one-to-one into the data lake (cf. Figure 1). As in our case, the LIMS is based on a large number of connected tables, a single, pre-generated "view" is mirrored on the data lake. Alternative ways consist in using existing export interfaces like HL7 or proprietary file formats. The direct access to the source system database facilitates data profiling, automatic identification of outliers and scaffolding the data catalogue for documenting the data, if such a repository is available at all. Nevertheless, an appalling issue can be the data structure in the source systems, which is rarely optimized to be queried via a CDWH. Data quality issues are handled in cooperation with the source system team or flagged as poor quality in order to filter these parts of the data out subsequently. The approach should be recursive: cleaning the sources as well as possible, tag

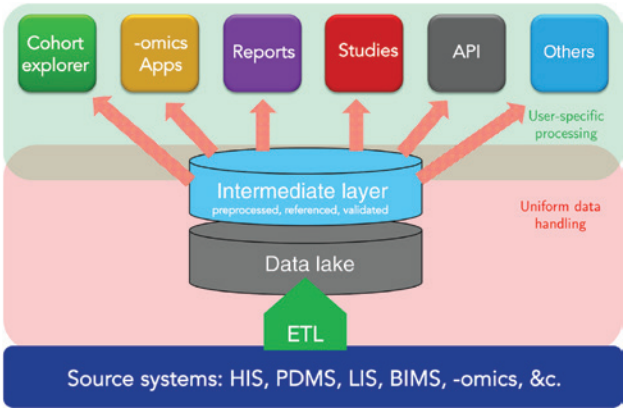

Figure 1: Simplified schematic of the Insel clinical data warehouse structure.
From bottom to top: Source systems (Hospital Information System [HIS], Patient Data Management System [PDMS], Laboratory Information System [LIS], Biobank Information Management System [BIMS], "-omics" systems [actually mainly file-based] and further source systems) are attached to the data lake via ETL (extract, transform, load) processes. The data is then preprocessed, referenced and at least partly validated and accessible in an intermediate layer. User-specific data transfers are based on the intermediate layer data and accessible via different tools or APIs, e.g. for queries from the SPHN.

problematic entries, elucidate the underlying problem and avoid re-occurrence within the sources. Especially in the LIMS, many auxiliary variables and quality control measurements are recorded, which are not relevant for a specific reported result. On the other hand, such data can be useful for "cockpit" applications or other monitoring tools. The use cases then define the processing of raw data to subsequent layers of the data warehouse according to "The Wisdom of Late Binding" postulated by Health Catalyst,[15] keeping in mind that other, not readily bound data fields of the source data system might be relevant for other use cases. Late binding allows to quickly adapt to new studies or use cases and to deliver the relevant dataset into a use case specific data mart, which at the Inselspital is entitled "Data Atelier". Data ateliers are – depending on the data governance – either identified, de-identified or fully anonymized. If the atelier owners need further adjustment, processing or formatting of the data, the delivery scripts can easily be adapted. The data pipelines that can be built to transform raw data into use case specific data marts vary from simple ETL processes or SQL queries to complex statistical procedures and modeling approaches or NLP algorithms for text mining.

15 https://www.healthcatalyst.com/late-binding-data-warehouse-explained/.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
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

1. Dixit R, Rogith D, Narayana V, Salimi M, Gururaj A, Ohno-Machado L, et al. User needs analysis and usability assessment of DataMed – a biomedical data discovery index. J Am Med Inform Assoc 2017;25:337–44.
2. Rodziewicz TL, Hipskind JE. Medical error prevention. Treasure Island (FL): StatPearls Publishing, 2018.
3. Leichtle AB, Dufour J-F, Fiedler GM. Potentials and pitfalls of clinical peptidomics and metabolomics. Swiss Med Wkly 2013;143:w13801.
4. Cadamuro J, Mrazek C, Leichtle AB, Kipman U, Felder TK, Wiedemann H, et al. Influence of centrifugation conditions on the results of 77 routine clinical chemistry analytes using standard vacuum blood collection tubes and the new BD-Barricor tubes. Biochem Med (Zagreb) 2017;28:764–10.
5. Liu X, Hoene M, Wang X, Yin P, Häring H-U, Xu G, et al. Serum or plasma, what is the difference? Investigations to facilitate the sample material selection decision making process for metabolomics studies and beyond. Anal Chim Acta 2018 (in press).
6. Cadamuro J, Gaksch M, Mrazek C, Haschke-Becher E, Plebani M. How do we use the data from pre-analytical quality indicators and how should we? J Lab Precis Med 2018;3:40.
7. Walz B, Fierz W. Der Referenzbereich ist tot – es lebe der Reference Change Value. Ther Umsch 2015;72:130–5.
8. Park J, Lee S, Kim Y, Choi A, Lee H, Lim J, et al. Comparison of four automated carcinoembryonic antigen immunoassays: ADVIA Centaur XP, ARCHITECT I2000sr, Elecsys E170, and Unicel Dxi800. Ann Lab Med 2018;38:355–61.
9. Bazsó FL, Ozohanics O, Schlosser G, Ludányi K, Vékey K, Drahos L. Quantitative comparison of tandem mass spectra obtained on various instruments. J Am Soc Mass Spectrom 2016;27:1357–65.
10. Bietenbeck A. Combining medical measurements from diverse sources: experiences from clinical chemistry. Stud Health Technol Inform 2016;228:58–62.
11. Master SR, Mayer-Schonberger V. Learning from our mistakes: the future of validating complex diagnostics. Clin Chem 2015;61:347–8.
12. Betsou F, Bilbao R, Case J, Chuaqui R, Clements JA, De Souza Y, et al. Standard PREanalytical code version 3.0. Biopreserv Biobank 2018;16:9–12.
13. Winter C, Ganslandt T, Bietenbeck A. No mathematical shortcuts for standardization or harmonization of laboratory measurements. Lab Medizin 2018;42:59–62.

14. Shi L, Li S, Yang X, Qi J, Pan G, Zhou B. Semantic health knowledge graph: semantic integration of heterogeneous medical knowledge and services. Biomed Res Int 2017;2017:1–12.
15. Cimino JJ, Hayamizu TF, Bodenreider O, Davis B, Stafford GA, Ringwald M. The caBIG terminology review process. J Biomed Inform 2009;42:571–80.
16. Deckard J, McDonald CJ, Vreeman DJ. Supporting interoperability of genetic data with LOINC. J Am Med Inform Assoc 2015;22:621–7.
17. Srivastava S. Informatics in proteomics. Boca Raton: CRC Press, 2005.
18. Tenenbaum JD, Blach C. Best practices and lessons learned from reuse of 4 patient-derived metabolomics datasets in Alzheimer's disease. Pac Symp Biocomput 2018;23:280–91.
19. Campbell JR, Talmon G, Cushman-Vokoun A, Karlsson D, Scott Campbell W. An extended SNOMED CT concept model for observations in molecular genetics. AMIA Annu Symp Proc 2016;2016:352–60.
20. Bloomfield RA Jr, Polo-Wood F, Mandel JC, Mandl KD. Opening the duke electronic health record to apps: implementing SMART on FHIR. Int J Med Inform 2017;99:1–10.
21. Lovis C. Digital health: a science at crossroads. Int J Med Inform 2018;110:108–10.
22. Murtagh MJ, Turner A, Minion JT, Fay M, Burton PR. International data sharing in practice: new technologies meet old governance. Biopreserv Biobank 2016;14:231–40.
23. Nind T, Galloway J, McAllister G, Scobbie D, Bonney W, Hall C, et al. The Research Data Management Platform (RDMP): a novel, process driven, open-source tool for the management of longitudinal cohorts of clinical data. GigaScience 2018;7:1–12.